Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Wed Jun 24 17:28:39 1998; MasPar time 3146.28 Seconds
1361.982 Million cell updates/sec
ular output not generated.

Title: >US-08-866-354-1
Description: (1-2568) from US08866354.seq
Perfect Score: 2525
N.A. Sequence: 1 AATCTGGCCTCGGAACACGC.....CACTCTTGTACAAAAGAAAA 2568
Comp: TTAGACCGGAGCCTTGTGCG.....GTGAGAACATGTTTTCTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 457396 seqs, 834342348 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl54
1:em_ba 2:em_htg 3:em_hum1 4:em_hum2 5:em_in 6:em_om
7:em_or 8:em_ov 9:em_pat 10:em_pl 11:em_ro 12:em_vi
Database: genbank106
13:gb_ba 14:gb_htg 15:gb_in 16:gb_om 17:gb_ov 18:gb_pat
19:gb_ph 20:gb_pl 21:gb_prl 22:gb_pr2 23:gb_ro 24:gb_st
25:gb_sy 26:gb_un 27:gb_vi

Statistics: Mean 12.226; Variance 7.675; scale 1.593

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	2285	90.5	2560	22	AF038421	Homo sapiens GPI-linke	0.00e+00
2	2011	79.6	2175	22	HSU95847	Human GDNF receptor al	0.00e+00
3	1619	64.1	1619	22	AF042080	Homo sapiens glial cel	0.00e+00
4	1571	62.2	1707	22	HSU97144	Homo sapiens RET ligan	0.00e+00
5	1243	49.2	3616	23	RNU97142	Rattus norvegicus RET	0.00e+00
6	1241	49.1	2138	23	RNU59486	Rattus norvegicus GDNF	0.00e+00
7	1223	48.4	2549	23	AF014117	Mus musculus GDNF rece	0.00e+00
8	1069	42.3	1651	23	AF015172	Mus musculus GDNF rece	0.00e+00
9	1061	42.0	1415	23	AB000800	Mouse mRNA for GDNF re	0.00e+00
10	970	38.4	1392	23	RN2072	Rattus norvegicus mRNA	0.00e+00
11	615	24.4	3037	17	GGU90541	Gallus gallus GDNF rec	0.00e+00
12	491	19.4	832	22	HSGFRA1G11	Homo sapiens GPI-linke	2.25e-297
13	488	19.3	777	23	AF012811	Mus musculus GDNF rece	2.39e-295
14	339	13.4	596	22	HSGFRA1G06	Homo sapiens GPI-linke	3.42e-195
15	298	11.8	385	22	HSGFRA1G01	Homo sapiens GPI-linke	7.33e-168
16	296	11.7	497	22	HSGFRA1G03	Homo sapiens GPI-linke	1.57e-166
17	291	11.5	529	22	HSGFRA1G02	Homo sapiens GPI-linke	3.29e-163
18	214	8.5	1392	23	AF002701	Mus musculus GDNF fami	2.11e-112
19	214	8.5	1395	22	HSU93703	Human glial cell line-	2.11e-112
20	214	8.5	1526	22	AF002700	Homo sapiens GDNF fami	2.11e-112
21	214	8.5	2906	22	HSU97145	Homo sapiens RET ligan	2.11e-112
22	210	8.3	2787	23	RNU97143	Rattus norvegicus RET	8.71e-110
23	208	8.2	1395	23	AF005226	Rattus norvegicus glia	1.77e-108
24	183	7.2	432	22	HSGFRA1G09	Homo sapiens GPI-linke	3.29e-92
25	166	6.6	2933	17	GGU90542	Gallus gallus neurturi	3.15e-81
26	138	5.5	397	22	HSGFRA1G08	Homo sapiens GPI-linke	2.47e-63
27	114	4.5	436	22	HSGFRA1G07	Homo sapiens GPI-linke	2.95e-48
28	96	3.8	385	18	A42902	Sequence 34 from Paten	3.57e-37
29	85	3.4	363	22	HSGFRA1G04	Homo sapiens GPI-linke	1.56e-30
c 30	63	2.5	7218	18	I66494	Sequence 14 from paten	1.08e-17
31	58	2.3	256	22	HSGFRA1G10	Homo sapiens GPI-linke	6.87e-15
32	50	2.0	7218	18	I66494	Sequence 14 from paten	1.58e-10
33	33	1.3	215	18	I28278	Sequence 5 from patent	4.93e-02
c 34	33	1.3	215	18	I28278	Sequence 5 from patent	4.93e-02
35	32	1.3	1203	22	AF051767	Homo sapiens GDNF fami	1.40e-01
36	32	1.3	4544	15	PFAARPS2	P.falciparum mRNA for	1.40e-01
37	34	1.3	10772	15	AF012089	Drosophila melanogaste	1.72e-02
c 38	33	1.3	10772	15	AF012089	Drosophila melanogaste	4.93e-02
39	30	1.2	1996	20	KCU53121	Komma caudata nucleomo	1.06e+00
40	30	1.2	7176	15	PFACPSI	Plasmodium falciparum	1.06e+00
c 41	30	1.2	137785	22	HS345P10	Human DNA sequence fro	1.06e+00
42	29	1.1	598	18	E02043	5' untranslated region	2.85e+00
43	29	1.1	1021	27	ORCPV160	Cowpox virus (CPV) 160	2.85e+00
44	29	1.1	4644	27	CPVATI	Cowpox virus A-type in	2.85e+00
45	29	1.1	4892	22	HUMSUR1RNA	Homo sapiens sulfonylu	2.85e+00

ALIGNMENTS

RESULT 1
LOCUS AF038421 2560 bp mRNA PRI 10-APR-1998
DEFINITION Homo sapiens GPI-linked anchor protein (GFRA1) mRNA, complete cds.
ACCESSION AF038421
NID g2921531
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2560)
AUTHORS Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S., Halushka,M., Fox,G.M. and Chakravarti,A.
TITLE Human GFRA1: Cloning, mapping, genomic structure, and evaluation as a candidate gene for Hirschsprung disease susceptibility
JOURNAL Genomics 48, 354-362 (1998)
REFERENCE 2 (bases 1 to 2560)
AUTHORS Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S., Halushka,M., Fox,G.M. and Chakravarti,A.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109 Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES Location/Qualifiers
source 1. .2560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25"
gene <1. .>2560
/gene="GFRA1"
CDS 557. .1954
/gene="GFRA1"
/note="similar to GDNF Receptor alpha; part of functional ligand for RET receptor tyrosine kinase"
/codon_start=1
/product="GPI-linked anchor protein"
/db_xref="PID:g2921532"
/translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI

```
                    YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKA
                    CNLDDICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                    IACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSS
                    CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKD
                    NTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTALRVKNKPLGPAGSENEIPTHVL
                    PPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVL
                    VVTALSTLLSLTETS"
BASE COUNT      642 a    675 c    632 g    611 t
ORIGIN

  Query Match          90.5%;  Score 2285;  DB 22;  Length 2560;
  Best Local Similarity  98.3%;  Pred. No. 0.00e+00;
  Matches  2525;  Conservative    0;  Mismatches   3;  Indels  40;  Gaps   2;

Db      18 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGCTTCCAATAACCACTAACATCCCTAA 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGCTTCCAATAACCACTAACATCCCTAA 60

        78 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGAG 137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        61 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGAG 120

Db     138 CTCTCGAAGATTACCGCATCTATTTTTTTTTTCTTTTTTTTCTTTTCCTAGCGCAGATAA 197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 CTCTCGAAGATTACCGCATCTATTTTTTTTTTCTTTTTTTTCTTTTCCTAGCGCAGATAA 180

Db     198 AGTGAGCCCGGAAAGGGAAGGAGGGGGCGGGGACACCATTGCCCTGAAAGAATAAATAAG 257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 AGTGAGCCCGGAAAGGGAAGGAGGGGGCGGGGACACCATTGCCCTGAAAGAATAAATAAG 240

Db     258 TAAATAAACAAACTGGCTCCTCGCCGCAGCTGGACGCGGTCGGTTGAGTCCAGGTTGGGT 317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 TAAATAAACAAACTGGCTCCTCGCCGCAGCTGGACGCGGTCGGTTGAGTCCAGGTTGGGT 300

Db     318 CGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATCCCGGAGCTGAGTCG 377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     301 CGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATCCCGGAGCTGAGTCG 360

Db     378 CCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGAGCTGAACT 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     361 CCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGAGCTGAACT 420

Db     438 TTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAA 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     421 TTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAA 480

Db     498 GACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCA 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     481 GACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCA 540

Db     558 TGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAG 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     541 TGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAG 600

Db     618 TGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGA 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     601 TGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGA 660

Db     678 GCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCA 737
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     661 GCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCA 720

Db     738 GCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGC 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     721 GCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGC 780

Db     798 AGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGC 857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     781 AGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGC 840

Db     858 GCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCAT 917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     841 GCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCAT 900

Db     918 ATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATG 977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     901 ATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATG 960

Db     978 TTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCT 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     961 TTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCT 1020

Db    1038 GCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCA 1097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1021 GCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCA 1080

Db    1098 GCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTG 1157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1081 GCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTG 1140

Db    1158 ACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCT 1217
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1141 ACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCT 1200

Db    1218 GCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGC 1277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1201 GCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGC 1260

Db    1278 CCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTG 1337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1261 CCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTG 1320

Db    1338 CGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAA 1397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1321 CGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAA 1380

Db    1398 ACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACT 1457
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1381 ACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACT 1440

Db    1458 ACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACG 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1441 ACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACG 1500

Db    1518 ACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATG 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1501 ACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATG 1560

Db    1578 CAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTAC 1637
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1561 CAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTAC 1620

Db    1638 AGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAG 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1621 AGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAG 1680

Db    1698 CAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCAC 1757
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1681 CAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCAC 1740

Db    1758 AGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATG 1817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1741 AGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATG 1800

Db    1818 AAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAA 1877
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1801 AAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAA 1860

Db    1878 GCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTT 1937
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1861 GCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTT 1920

Db    1938 TAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGACAAAAA 1997
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1921 TAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGACAAAAA 1980
```

```
Db   1998 CCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTCAGTTG 2057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1981 CCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTCAGTTG 2040

Db   2058 AGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTT-CCTTTTAAGAAAGCTTCTTGTG 2116
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Qy   2041 AGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTTTTAAGAAAGCTTCTTGTG 2100

Db   2117 ATCCTTCGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTCAGAAGGCTTT 2176
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2101 ATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTCAGAAGGCTTT 2160

Db   2177 GGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCAAACTGGGGCT 2236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2161 GGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCAAACTGGGGCT 2220

Db   2237 GTGTTTTCGATGATGATGATCATCATGATCATGAT------------------------ 2271
          |||||||||||||||||| |||||||| |||||
Qy   2221 GTGTTTTCGATGATGATGATNATCATGATNATGATNNNNNNNNNNNNNNNNNNNNNNNNN 2280

Db   2272 -------------GATTTTAACAGTTTTACTTCTGGCCTTTCCTAGCTAGAGAAGGAGT 2317
                        ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2281 NNNNNNNNNNNNNGATTTTAACAGTTTTACTTCTGGCCTTTCCTAGCTAGAGAAGGAGT 2340

Db   2318 TAATATTTCTAAGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTAATGATATAAA 2377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2341 TAATATTTCTAAGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTAATGATATAAA 2400

Db   2378 TTTCAGCCTACATTGATGCCAAGCTTTTTTGCCACAAAGAAGATTCTTACCAAGAGTGGG 2437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2401 TTTCAGCCTACATTGATGCCAAGCTTTTTTGCCACAAAGAAGATTCTTACCAAGAGTGGG 2460

Db   2438 CTTTGTGGAAACAGCTGGTACTGATGTTCACCTTTATATATGTACTAGCATTTTCCACGC 2497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2461 CTTTGTGGAAACAGCTGGTACTGATGTTCACCTTTATATATGTACTAGCATTTTCCACGC 2520

Db   2498 TGATGTTTATGTACTGTAAACAGTTCTGCACTCTTGTACAAAAGAAAA 2545
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2521 TGATGTTTATGTACTGTAAACAGTTCTGCACTCTTGTACAAAAGAAAA 2568


RESULT    2
LOCUS       HSU95847     2175 bp    mRNA            PRI       02-OCT-1997
DEFINITION  Human GDNF receptor alpha mRNA, complete cds.
ACCESSION   U95847
NID         g2459741
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2175)
  AUTHORS   Hishiki,T., Kondoh,K., Ichimiya,S., Nimura,Y., Seki,N., Ozaki,T.,
            Sakiyama,S., Takahashi,H., Ohnuma,N., Tanabe,M., Fujimura,S. and
            Nakagawara,A.
  TITLE     GDNF-induced differentiation and its enhancement by retinoic acid
            in primary human neuroblastomas expressing c-Ret and GDNFR-alpha
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2175)
  AUTHORS   Nakagawara,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1997) Division of Biochemistry, Chiba Cancer
            Center Research Institute, 666-2, Nitona, Chuoh-ku, Chiba 260,
            Japan
FEATURES             Location/Qualifiers
     source          1. .2175
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /tissue_type="substantia nigra"
     CDS             245. .1627
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2459742"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISVEHIPKGNNCLDAAKACNLDD
                     ICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                     RRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKEN
                     YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTAL
                     STLLSLTETS"
BASE COUNT      544 a    566 c    535 g    516 t     14 others
ORIGIN

  Query Match           79.6%;  Score 2011;  DB 22;  Length 2175;
  Best Local Similarity  96.5%;  Pred. No. 0.00e+00;
  Matches  2114;  Conservative    0;  Mismatches  61;  Indels  15;  Gaps   1;

Db      1 TGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATCCCGGAGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    296 TGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATCCCGGAGCTG 355

Db     61 AGTCGCCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGAGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    356 AGTCGCCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGAGCT 415

Db    121 GAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCATGCTGAGCTCCCTC 180
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Qy    416 GAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTC 475

Db    181 GGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    476 GGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGG 535

Db    241 CACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    536 CACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGC 595

Db    301 CGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    596 CGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGA 655

Db    361 GCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    656 GCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAA 715

Db    421 CTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    716 CTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCT 775

Db    481 GAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    776 GAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTG 835

Db    541 CCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    836 CCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTC 895

Db    601 CCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    896 CCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATC 955

Db    661 AG---------------TGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAA 705
          ||               ||||||||||||||||||||||||||||||||||||||||||||
Qy    956 AGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAA 1015

Db    706 GGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCAC 765
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1016 GGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCAC 1075

Db    766 CACCAGCGTGTCCAATGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTT 825
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Qy   1076 CACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTT 1135
```

```
Db    826 CTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACAT 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1136 CTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACAT 1195

Db    886 CGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGA 945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1196 CGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGA 1255

Db    946 GAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCG 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1256 GAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCG 1315

Db   1006 CCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAA 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1316 CCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAA 1375

Db   1066 GGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCC 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1376 GGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCC 1435

     1126 CAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1436 CAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGG 1495

Db   1186 GAACGACCTAGAAGAGTGCTTGAAATTCTTGAATTTCTTCAAGGACAATACATGTCTTAA 1245
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Qy   1496 GAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAA 1555

Db   1246 AAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCC 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1556 AAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCC 1615

Db   1306 AGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAACCCCTGGG 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy   1616 AGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGG 1675

Db   1366 GCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1676 GCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACA 1735

Db   1426 GGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAA 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1736 GGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAA 1795

Db   1486 TTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
~~   1796 TTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1855

     1546 TCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATT 1605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1856 TCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATT 1915

Db   1606 ATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGAC 1665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1916 ATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGAC 1975

Db   1666 AAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTC 1725
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1976 AAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTC 2035

Db   1726 AGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTTCCTTTTAAAAAAGCTTC 1785
          ||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Qy   2036 AGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTTNCCTTTTAAGAAAGCTTC 2095

Db   1786 TTGTGATCCTTTGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTCAAAAG 1845
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Qy   2096 TTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTCAGAAG 2155

Db   1846 GCTTNGGGATATGCTGTNTTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAANCAAACTG 1905
          |||| |||||||||||| |||||||||||||||||||||||||||||||||| |||||||
Qy   2156 GCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCAAACTG 2215

Db   1906 GGGNTGTGTTTTCGATGATGATGATGATCATGATGATGATCCTCGTGATCATGANGATGA 1965
          ||| |||||||||||||||||||| |||||||| |||||           |
Qy   2216 GGGCTGTGTTTTCGATGATGATGATNATCATGATNATGATNNNNNNNNNNNNNNNNNNNN 2275

Db   1966 TCNTCATGATCATGATNATGATTTTAACAGTTTTACTTCTGGCCTTTCCTANCTANANAA 2025
            |             | ||||||||||||||||||||||||||||||||||| ||| | ||
Qy   2276 NNNNNNNNNNNNNNNNNNNGATTTTAACAGTTTTACTTCTGGCCTTTCCTAGCTAGAGAA 2335

Db   2026 NGAGTTAATATTTCTANGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTAATGAT 2085
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Qy   2336 GGAGTTAATATTTCTAAGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTAATGAT 2395

Db   2086 ATAAATTTCACCCTACATTGATGCCAAGCTTTTTTGCCACAAAAAAAATTCTTNCCAAAA 2145
          |||||||||| |||||||||||||||||||||||||||||||| || |||||| |||| |
Qy   2396 ATAAATTTCAGCCTACATTGATGCCAAGCTTTTTTGCCACAAAGAAGATTCTTACCAAGA 2455

Db   2146 NTGGGCTTTGTGGAAACAGCTGGTACTGAT 2175
          |||||||||||||||||||||||||||||
Qy   2456 GTGGGCTTTGTGGAAACAGCTGGTACTGAT 2485
```

```
RESULT    3
LOCUS       AF042080     1619 bp    mRNA            PRI       22-JAN-1998
DEFINITION  Homo sapiens glial cell line-derived neurotrophic factor receptor
            alpha (GFRA1) mRNA, complete cds.
ACCESSION   AF042080
NID         g2801556
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1619)
  AUTHORS   Shefelbine,S.E., Khorana,S., Schultz,P.N., Huang,E., Thobe,N.,
            Hu,Z.J., Fox,G.M., Jing,S., Cote,G.J. and Gagel,R.F.
  TITLE     Mutational analysis of the GDNF/RET-GDNFRa signaling complex in a
            kindred with vesicoureteral reflux
  JOURNAL   Hum. Genet. (1998) In press
REFERENCE   2  (bases 1 to 1619)
  AUTHORS   Shefelbine,S.E., Khorana,S., Schultz,P.N., Huang,E., Thobe,N.,
            Hu,Z.J., Fox,G.M., Jing,S., Cote,G.J. and Gagel,R.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1998) Endocrinology-Box 15, M.D. Anderson Cancer
            Center, 1515 Holcombe Blvd, Houston, TX 77030, USA
FEATURES             Location/Qualifiers
     source          1. .1619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25-q26"
                     /tissue_type="medullary thyroid carcinoma tumor"
     gene            1. .1619
                     /note="GDNFRa"
                     /gene="GFRA1"
     CDS             111. .1508
                     /gene="GFRA1"
                     /note="GPI-linked receptor"
                     /codon_start=1
                     /product="glial cell line-derived neurotrophic factor
                     receptor alpha"
                     /db_xref="PID:g2801557"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKA
                     CNLDDICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     IACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSS
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVL
                     VVTALSTLLSLTETS"
     misc_feature    529. .543
                     /gene="GFRA1"
                     /note="found to be alternative exon after partial analysis
                     of genomic sequence"
```

```
BASE COUNT      409 a    448 c    419 g    343 t
ORIGIN

  Query Match           64.1%;  Score 1619;  DB 22;  Length 1619;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  1619;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 CAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAAGACCCAGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    430 CAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAAGACCCAGCG 489

Db     61 GCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCATGTTCCTGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    490 GCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCATGTTCCTGG 549

Db    121 CGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAGTGAGCGGCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    550 CGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAGTGAGCGGCG 609

Db    181 GAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTGCAGCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    610 GAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTGCAGCA 669

Db    241 CCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCTGGCAT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    670 CCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCTGGCAT 729

Db    301 CCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTCGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    730 CCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTCGC 789

Db    361 TCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTACT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    790 TCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTACT 849

Db    421 GGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGAACCAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    850 GGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGAACCAG 909

Db    481 TTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTTTCAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    910 TTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTTTCAGC 969

Db    541 AAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    970 AAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCG 1029

Db    601 ACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1030 ACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCA 1089

Db    661 ACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1090 ACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCC 1149

Db    721 CGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1150 CGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGC 1209

Db    781 GGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1210 GGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTT 1269

Db    841 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1270 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 1329

Db    901 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1330 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 1389

Db    961 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1390 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1449

Db    1021 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1450 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1509

Db    1081 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1510 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1569

Db    1141 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1570 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1629

Db    1201 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1630 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1689

Db    1261 AGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1690 AGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGA 1749

Db    1321 AATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAG 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1750 AATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAG 1809

Db    1381 GTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTGGTC 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1810 GTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTGGTC 1869

Db    1441 TGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACAGAAA 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1870 TGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACAGAAA 1929

Db    1501 CATCATAGCTGCATTAAAAAAAATACAATATGGACATGTAAAAAGACAAAAACCAAGTTAT 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1930 CATCATAGCTGCATTAAAAAAAATACAATATGGACATGTAAAAAGACAAAAACCAAGTTAT 1989

Db    1561 CTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTCAGTTGAGAAACAG 1619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1990 CTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGGAGCTCAGTTGAGAAACAG 2048


RESULT    4
LOCUS       HSU97144     1707 bp    mRNA            PRI       29-JUL-1997
DEFINITION  Homo sapiens RET ligand 1 (RETL1) mRNA, complete cds.
ACCESSION   U97144
NID         g2282025
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 1707)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Glial cell line-derived neurotrophic factor-dependent RET
            activation can be mediated by two different cell-surface accessory
            proteins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6238-6243 (1997)
  MEDLINE   97322356
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1997) Molecular Genetics, BIOGEN, 14 Cambridge
            Center, Cambridge, MA 02142, USA
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /dev_stage="embryonic"
                     /tissue_type="kidney"
     gene            1. .1707
                     /gene="RETL1"
     CDS             262. .1644
                     /gene="RETL1"
                     /note="PI-linked cell-surface accessory protein; GDNFRa"
                     /codon_start=1
                     /product="RET ligand 1"
                     /db_xref="PID:g2282026"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISVEHIPKGNNCLDAAKACNLDD
                     ICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                     RRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKEN
                     YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTAL
                     STLLSLTETS"
BASE COUNT      420 a    490 c    456 g    341 t
ORIGIN

  Query Match           62.2%;  Score 1571;  DB 22;  Length 1707;
  Best Local Similarity  98.5%;  Pred. No. 0.00e+00;
  Matches  1676;  Conservative    0;  Mismatches   9;  Indels  16;  Gaps   2;

Db      2 CGGCCGCGTCGACGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTC 61
          |||  | |||  | ||||||||||||||||||||||||||||||||||||||||||||
Qy    281 CGGTTGAGTCCAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTC 340

Db     62 GCCATCCCGGAGCTGAGTCGCCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTT 121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    341 GCCATCCCGGAGCTGAGTCGCCGGCGGCGGTGGCTGCTGCCAGACCCGGAGTTTCCTCTT 400

Db    122 TCACTGGATGGAGCTGAACTTTGGGCGGCCAGAACAGCACAGCTGTCCGGGGATCGCTGC 181
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Qy    401 TCACTGGATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGC 460

Db    182 ACGCTGAGCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTTGGGGGGGCGGGG 241
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Qy    461 ACGCTGAGCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTT-GGGGGGGCGGGG 519

Db    242 ACCAGCCCCGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGG 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    520 ACCAGCCCCGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGG 579

Db    302 ACTTGCTCCTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTG 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    580 ACTTGCTCCTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTG 639

Db    362 ATCAGTGCCTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGG 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    640 ATCAGTGCCTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGG 699

Db    422 CGGGCAAGGAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCA 481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    700 CGGGCAAGGAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCA 759

Db    482 GCGCCATGGAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGA 541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    760 GCGCCATGGAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGA 819

Db    542 AGAAGGAGAAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATG 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    820 AGAAGGAGAAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATG 879

Db    602 ATCTGCTGGAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGG 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    880 ATCTGCTGGAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGG 939

Db    662 TGGTCCCATTCATATCAG----------------TGGAGCACATTCCCAAAGGGAACAACT 706
          |||||||||||||||||                ||||||||||||||||||||||||||||

Qy    940 TGGTCCCATTCATATCAGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACT 999

Db    707 GCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGT 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1000 GCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGT 1059

Db    767 ACATCACCCCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCGAAGTGCCACA 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1060 ACATCACCCCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCGAAGTGCCACA 1119

Db    827 AGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCT 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1120 AGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCT 1179

Db    887 GCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCT 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1180 GCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCT 1239

Db    947 CCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATT 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1240 CCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATT 1299

Db   1007 ACATCTGCAGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTG 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1300 ACATCTGCAGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTG 1359

Db   1067 TCAGCAGCTGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTG 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1360 TCAGCAGCTGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTG 1419

Db   1127 GCACAGTCATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTG 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1420 GCACAGTCATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTG 1479

Db   1187 ACTGCAGCAACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGG 1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1480 ACTGCAGCAACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGG 1539

Db   1247 ACAATACATGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGT 1306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1540 ACAATACATGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGT 1599

Db   1307 GGCAGCCAGCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTA 1366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1600 GGCAGCCAGCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTA 1659

Db   1367 AGAACAAGCCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCAC 1426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1660 AGAACAAGCCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCAC 1719

Db   1427 CGTGTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCT 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1720 CGTGTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCT 1779

Db   1487 GTATTTCCAATGGTAATTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAA 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1780 GTATTTCCAATGGTAATTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAA 1839

Db   1547 AATCAATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCG 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1840 AATCAATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCG 1899

Db   1607 CTCTGTCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATAT 1666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1900 CTCTGTCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATAT 1959

Db   1667 GGACATGAAAAAAAAAAAAAA 1687
          ||||||| ||||| | |||||
Qy   1960 GGACATGTAAAAAGACAAAAA 1980


RESULT    5
LOCUS       RNU97142     3616 bp    mRNA            ROD       29-JUL-1997
```

```
DEFINITION  Rattus norvegicus RET ligand 1 (RETL1) mRNA, complete cds.
ACCESSION   U97142
NID         g2282021
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 3616)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Glial cell line-derived neurotrophic factor-dependent RET
            activation can be mediated by two different cell-surface accessory
            proteins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6238-6243 (1997)
  MEDLINE   97322356
REFERENCE   2  (bases 1 to 3616)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1997) Molecular Genetics, BIOGEN, 14 Cambridge
            Center, Cambridge, MA 02142, USA
FEATURES             Location/Qualifiers
     source          1. .3616
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /dev_stage="embryonic day 18"
                     /tissue_type="kidney"
     gene            1. .3616
                     /gene="RETL1"
     CDS             257. .1663
                     /gene="RETL1"
                     /note="PI-linked cell-surface accessory protein; GDNF
                     receptor alpha"
                     /codon_start=1
                     /product="RET ligand 1"
                     /db_xref="PID:g2282022"
                     /translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     IACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPV
                     LMLTALAALLSVSLAETS"
BASE COUNT      876 a    955 c    921 g    864 t
ORIGIN
```

Query Match 49.2%; Score 1243; DB 23; Length 3616;
Best Local Similarity 83.9%; Pred. No. 0.00e+00;
Matches 1657; Conservative 0; Mismatches 297; Indels 20; Gaps 13;

```
Db      8 CAGGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCC 67
          |||||||||||||| |||||||||| |||||||  ||||||||||||||||  |  |||
Qy    291 CAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGC--CATCCC 348

Db     68 GGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGG 127
          ||| |||||||| |||||||||| | || |   |||  ||||| | ||||  | ||||
Qy    349 GGAGCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGG 407

Db    128 ATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGA 187
          |||||||||||||||| | |||||||| ||| |||  | ||||||||||||||||||||
Qy    408 ATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGA 467

Db    188 GCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTG 244
          |||| |||  | |||||  |||||||||  |||||||   ||||||||||||||||||
Qy    468 GCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCC 527

Db    245 CGCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTG 304
          |||| |||||||||||||| || || |||||||||||||||| ||| |||| |||||
Qy    528 CGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTC 587

Db    305 ATGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGC 364
          |||| ||||| ||||| || |||||||| ||||| || |||||||||| |||||||||
Qy    588 CTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGC 647

Db    365 CTGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAG 424
          |||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||
Qy    648 CTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAG 707

Db    425 GAAACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATG 484
          || |||||||||||||| |||||||||| |||||||||||||||||||| |||||||||
Qy    708 GAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATG 767

Db    485 GAGGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAG 544
          |||||| |||||||||||||| || |||||||||||||||||||||| |||||||| |||
Qy    768 GAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAG 827

Db    545 AAGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTG 604
          ||||| || ||||| || |||||||||||||||||||||||||||||||||| || |||
Qy    828 AAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTG 887

Db    605 GAAGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCG 664
          || |||||||| ||||| || |||||||||| |||||||||||||||||| |||||
Qy    888 GAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCA 947

Db    665 TTCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGAC 724
          |||||||||||||||| |||||||||| |||||| ||||||||||||||||||||||
Qy    948 TTCATATCAGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGAT 1007

Db    725 GCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACC 784
          ||||| |||||||||||||| |||||||  || ||||||||||||||||| |||||||||
Qy   1008 GCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACC 1067

Db    785 CCCTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTC 844
          || |||||||||||| |||||||||| |||||||||||||||||| |||||||||||||
Qy   1068 CCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTC 1127

Db    845 AGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGC 904
          |||||||||| |||||||| |||||||||||||||||||| ||||||||||||||||||
Qy   1128 CGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGC 1187

Db    905 CGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAA 964
          |||||||||||||||| |||||| |||||||||| ||||| || ||||||||||||||||
Qy   1188 CGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAA 1247

Db    965 GAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGC 1024
          ||  |  |||| ||||||||||  ||| | |||| |||||||||||||| ||||||||||
Qy   1248 GAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGC 1307

Db   1025 AGATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAAC 1084
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||| |
Qy   1308 AGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGC 1367

Db   1085 TGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTC 1144
          ||||| ||||| |||||||| |||||||||| |||||||||| || ||||||||||||
Qy   1368 TGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTC 1427

Db   1145 ATGACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGC 1204
          ||||| |||||||| |||||||||| |||||||| ||||| ||||||||||||||||||
Qy   1428 ATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGC 1487

Db   1205 AACAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACT 1264
          ||||| || || ||||| ||||| |||||||||||| ||||||| || ||||||||||||
Qy   1488 AACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACA 1547

Db   1265 TGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCA 1324
          ||||| |||||||||||||||||||||||||||||||| ||||||||| ||||||||||
Qy   1548 TGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCA 1607

Db   1325 GCCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAG 1384
```

```
Qy   1608 GCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAG 1667

Db   1385 CCTCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCG 1444

Qy   1668 CCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCA 1727

Db   1445 AATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCT 1504

Qy   1728 AATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCC 1787

Db   1505 GATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCA 1564

Qy   1788 AATGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCA 1844

Db   1565 ATGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTT 1624

Qy   1845 ATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTG 1904

     1625 GCTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTAT 1684

Qy   1905 TCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACA 1964

Db   1685 GAAAAGACAAAAGAGAACCAAGTATTCTGTC-CCTGTCCTCTTGTATATCTGAAAATCCA 1743

Qy   1965 TGTAAAAAGACAAA-AACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCA 2023

Db   1744 GTTTTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAAC-TCTTTCCTTGT---TT 1799

Qy   2024 GTTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTT 2082

Db   1800 TTAAGAAAGCTTGT-G-GC-CCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATT 1856

Qy   2083 TTAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCAT 2142

Db   1857 CCAAACCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAA 1916

Qy   2143 CCAAACTCAGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCT 2202

Db   1917 GCAAGCTGGGCTTATCATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGAT 1970

Qy   2203 GTAAAGCAAACTGGGGCTGTGTTTTC-GATGATGATGATNATCATGATNATGAT 2255
```

```
RESULT    6
LOCUS       RNU59486     2138 bp    mRNA            ROD       23-JUL-1996
DEFINITION  Rattus norvegicus GDNF receptor alpha mRNA, complete cds.
ACCESSION   U59486
            g1399862
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 2138)
  AUTHORS   Jing,S.J., Wen,D., Yu,Y., Holst,P.L., Luo,Y., Fang,M., Tamir,R.,
            Antonio,L., Hu,Z., Cupples,R., Louis,J.-C., Hu,S., Altrock,B. and
            Fox,G.M.
  TITLE     GDNF-induced activation of the ret protein tyrosine kinase is
            mediated by GDNFR-alpha, a novel receptor for GDNF
  JOURNAL   Cell 85 (7), 1113-1124 (1996)
  MEDLINE   96270513
REFERENCE   2  (bases 1 to 2138)
  AUTHORS   Fox,G.M., Jing,S.J., Yu,Y., Holst,P.L., Fang,M., Tamir,R.,
            Antonio,L., Hu,Z. and Louis,J.-C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-1996) Immunology, Amgen, Inc., Amgen Center,
            Thousand Oaks, CA 91320-1789, USA
FEATURES             Location/Qualifiers
     source          1. .2138
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             302. .1708
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g1399863"
                     /translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     IACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPV
                     LMLTALAALLSVSLAETS"
BASE COUNT      494 a    593 c    584 g    467 t
ORIGIN
```

```
  Query Match           49.1%;  Score 1241;  DB 23;  Length 2138;
  Best Local Similarity  83.9%;  Pred. No. 0.00e+00;
  Matches  1656;  Conservative    0;  Mismatches 298;  Indels  20;  Gaps  13;

Db     53 CAAGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCC 112

Qy    291 CAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGC--CATCCC 348

Db    113 GGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGG 172

Qy    349 GGAGCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGG 407

Db    173 ATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGA 232

Qy    408 ATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGA 467

Db    233 GCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTG 289

Qy    468 GCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCC 527

Db    290 CGCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTG 349

Qy    528 CGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTC 587

Db    350 ATGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGC 409

Qy    588 CTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGC 647

Db    410 CTGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAG 469

Qy    648 CTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAG 707

Db    470 GAAACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATG 529

Qy    708 GAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATG 767

Db    530 GAGGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAG 589

Qy    768 GAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAG 827

Db    590 AAGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTG 649

Qy    828 AAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTG 887

Db    650 GAAGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCG 709

Qy    888 GAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCA 947

Db    710 TTCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGAC 769

Qy    948 TTCATATCAGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGAT 1007

Db    770 GCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACC 829

Qy   1008 GCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACC 1067

Db    830 CCCTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTC 889
```

```
Qy   1068 CCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTC 1127

Db    890 AGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGC 949

Qy   1128 CGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGC 1187

Db    950 CGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAA 1009

Qy   1188 CGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAA 1247

Db   1010 GAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGC 1069

Qy   1248 GAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGC 1307

Db   1070 AGATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAAC 1129

Qy   1308 AGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGC 1367

     1130 TGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTC 1189

     1368 TGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTC 1427

Db   1190 ATGACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGC 1249

Qy   1428 ATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGC 1487

Db   1250 AACAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACT 1309

Qy   1488 AACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACA 1547

Db   1310 TGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCA 1369

Qy   1548 TGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCA 1607

Db   1370 GCCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAG 1429

Qy   1608 GCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAG 1667

Db   1430 CCTCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCG 1489

Qy   1668 CCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCA 1727

Db   1490 AATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCT 1549

Qy   1728 AATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCC 1787

Db   1550 GATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCA 1609

Qy   1788 AATGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCA 1844

Db   1610 ATGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTT 1669

Qy   1845 ATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTG 1904

Db   1670 GCTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTAT 1729

Qy   1905 TCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACA 1964

Db   1730 GAAAAGACAAAAGAGAACCAAGTATTCTGTC-CCTGTCCTCTTGTATATCTGAAAATCCA 1788

Qy   1965 TGTAAAAAGACAAA-AACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCA 2023

Db   1789 GTTTTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAAC-TCTTTCCTTGT---TT 1844

Qy   2024 GTTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTT 2082

Db   1845 TTAAGAAAGCTTGT-G-GC-CCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATT 1901

Qy   2083 TTAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCAT 2142

Db   1902 CCAAACCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAA 1961

Qy   2143 CCAAACTCAGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCT 2202

Db   1962 GCAAGCTGGGCTTATCATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGAT 2015

Qy   2203 GTAAAGCAAACTGGGGCTGTGTTTTC-GATGATGATGATNATCATGATNATGAT 2255
```

```
RESULT    7
LOCUS       AF014117     2549 bp    mRNA            ROD       20-NOV-1997
DEFINITION  Mus musculus GDNF receptor alpha (GDNFR-alpha) mRNA, complete cds.
ACCESSION   AF014117
NID         g2624960
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2549)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Cloning of a novel murine isoform of the glial cell line-derived
            neurotrophic factor receptor
  JOURNAL   Neuroreport 9 (1) (1998) In press
REFERENCE   2  (bases 1 to 2549)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Biochemistry, 10 Kent Ridge Crescent,
            Singapore 119260, Singapore
FEATURES             Location/Qualifiers
     source          1. .2549
                     /organism="Mus musculus"
                     /strain="C57"
                     /db_xref="taxon:10090"
                     /tissue_type="liver"
     gene            1. .2549
                     /gene="GDNFR-alpha"
     CDS             914. .2320
                     /gene="GDNFR-alpha"
                     /note="glial cell-line derived neurotrophic growth factor
                     receptor"
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2624961"
                     /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     VACTERRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPV
                     MVFTALAALLSVSLAETS"
BASE COUNT      623 a    680 c    693 g    553 t
ORIGIN
```

```
  Query Match           48.4%;  Score 1223;  DB 23;  Length 2549;
  Best Local Similarity  85.1%;  Pred. No. 0.00e+00;
  Matches  1608;  Conservative    0;  Mismatches 260;  Indels  21;  Gaps  13;

Db    665 CAGGTTGGGTCGGAACTGAACCCCTAAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCTC 724

Qy    291 CAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATC-C-C 348

Db    725 GGAGCTGAGTCCCTGGCGGCGGTGGGCGGCAGAGCAACGGGGAGTCTGCTCTCACCCTGG 784

Qy    349 GGAGCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGG 407

Db    785 ATGGAGCTTAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCTCTGCACGCTGA 844

Qy    408 ATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGA 467

Db    845 GCTCTCTCCGCGAGATCCGGTGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTG 901

Qy    468 GCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTGGGGGGGCGGGGACCAGCCC 527
```

```
Db    902 CGCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTG 961
Qy    528 CGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTC 587

Db    962 ATGTCGGCCGAGGTGAGTGGTGGGGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGC 1021
Qy    588 CTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGC 647

Db   1022 CTGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAG 1081
Qy    648 CTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAG 707

Db   1082 GAAACCAACTTCAGCCTGACATCCGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATG 1141
Qy    708 GAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATG 767

Db   1142 GAGGCCTTGAAGCAGAAGTCTCTCTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAG 1201
Qy    768 GAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAG 827

Db   1202 AAGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTG 1261
Qy    828 AAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTG 887

Db   1262 GAAGATTCCCCATACGAGCCGGTTAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCG 1321
Qy    888 GAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCA 947

Db   1322 TTCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTCGAT 1381
Qy    948 TTCATATCAGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGAT 1007

Db   1382 GCAGCCAAGGCCTGCAACCTGGATGACACCTGCAAGAAGTACAGGTCCGCCTACATCACC 1441
Qy   1008 GCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACC 1067

Db   1442 CCCTGTACCACCAGCATGTCCAATGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTC 1501
Qy   1068 CCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTC 1127

Db   1502 AGGCAGTTCTTCGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGC 1561
Qy   1128 CGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGC 1187

Db   1562 CGGGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAA 1621
Qy   1188 CGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAA 1247

Db   1622 GAACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGC 1681
Qy   1248 GAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGC 1307

Db   1682 AGATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAAC 1741
Qy   1308 AGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGC 1367

Db   1742 TGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTC 1801
Qy   1368 TGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTC 1427

Db   1802 ATGACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGC 1861
Qy   1428 ATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGC 1487

Db   1862 AACAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACG 1921
Qy   1488 AACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACA 1547

Db   1922 TGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAACCA 1981
Qy   1548 TGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCA 1607

Db   1982 GCCCCCCCAGTCCAGACCACCACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAG 2041
Qy   1608 GCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAG 1667

Db   2042 CCTCTAGGGCCAGCAGGCTCTGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCT 2101
Qy   1668 CCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCA 1727

Db   2102 AATTTGCAGGCACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCT 2161
Qy   1728 AATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCC 1787

Db   2162 GATAATGATTACGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCA 2221
Qy   1788 AATGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCA 1844

Db   2222 ATGGCTGCTCCTCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTG 2281
Qy   1845 ATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTG 1904

Db   2282 GCTGCCCTGTTGTCTGTATCATTGGCAGAA-ACAT-CGT-AGCTGCATCCAGTATGAAAA 2338
Qy   1905 TCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACA 1964

Db   2339 -GACAAAAGA---AAACCAAGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAG 2394
Qy   1965 TGTAAAAAGACAAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAG 2024

Db   2395 TTTTAAAAGCTCGGTTGAGAAGCAGTTTCACCCAGCGGGAAC-TCCTTTGTTGT---TTT 2450
Qy   2025 TTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTTT 2083

Db   2451 TAAGAAAGCTTTTTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTC 2510
Qy   2084 TAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATC 2143

Db   2511 CAAACTCAGAAAACTCTGGGGCATGGTGT 2539
Qy   2144 CAAACTCAGAAGGCTTTGGGATATGCTGT 2172
```

```
RESULT    8
LOCUS       AF015172     1651 bp    mRNA            ROD       20-NOV-1997
DEFINITION  Mus musculus GDNF receptor beta (GDNFR-beta) mRNA, complete cds.
ACCESSION   AF015172
NID         g2624962
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Cloning of a novel murine isoform of the glial cell line-derived
            neurotrophic factor receptor
  JOURNAL   Neuroreport 9 (1) (1998) In press
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1997) Biochemistry, National University of
            Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore
FEATURES             Location/Qualifiers
     source          1. .1651
                     /organism="Mus musculus"
                     /strain="C57"
                     /db_xref="taxon:10090"
     gene            1. .1651
                     /gene="GDNFR-beta"
     CDS             230. .1621
                     /gene="GDNFR-beta"
                     /note="glial cell-line derived neurotrophic factor
                     receptor beta."
                     /codon_start=1
```

```
                  /product="GDNF receptor beta"
                  /db_xref="PID:g2624963"
                  /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                  KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                  YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISVEHISKGNNCLDAAKACNLDD
                  TCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTE
                  RRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKEN
                  YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLK
                  NAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCAN
                  LQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTA
                  LAALLSVSLAETS"
BASE COUNT      394 a    475 c    443 g    339 t
ORIGIN

  Query Match           42.3%;  Score 1069;  DB 23;  Length 1651;
  Best Local Similarity  86.7%;  Pred. No. 0.00e+00;
  Matches  1405;  Conservative    0;  Mismatches 192;  Indels  24;  Gaps  10;

Db      1 CCCCTAAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCTCGGAGCTGAGTCCCTGGCGGC 60
          ||||||||||||  |||||||||||||||||  | || |||||||||| | ||||||
Qy    311 CCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATC-C-CGGAGCTGAGTCGCCGGCGGC 368

Db     61 GGTGGGCGGCAGAGCAACGGGGAGTCTGCTCTCACCCTGGATGGAGCTTAACTTTGAGTG 120
          ||||| | || |    ||  ||||| | ||||  | ||||||||||| |||||||| | |
Qy    369 GGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGAGCTGAACTTTGGGCG 427

Db    121 GCCAGAGGAGCGCAGTCGCCCGGGGATCTCTGCACGCTGAGCTCTCTCCGCGAGATCCGG 180
          ||||||| ||| |||  | |||||||| |||||||||||||| ||| || ||| || |
Qy    428 GCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAAGACCCAG 487

Db    181 TGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTGCGCGGCGGCACCATGTTCCT 237
          |||||||  |||||||   |||||||||||||||||  ||||  ||||||||||||||||
Qy    488 CGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCATGTTCCT 547

Db    238 AGCCACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTGATGTCGGCCGAGGTGAGTGG 297
          || || |||||||||| |||||| ||| |||| |||||  |||||||||| ||||| ||
Qy    548 GGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAGTGAGCGG 607

Db    298 TGGGGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGCCTGAAGGAACAGAGCTGCAG 357
          || |||||||||| || ||||||||||||||||||||||||||||| ||||||||||
Qy    608 CGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTGCAG 667

Db    358 CACCAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTGAC 417
          ||||||||||||| || |||||||||||||||||||||||||||||| ||||||||||| |
Qy    668 CACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCTGGC 727

Db    418 ATCCGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATGGAGGCCTTGAAGCAGAAGTC 477
          |||||||| |||||||||||||||||||||||||||||| |||||||| ||||||||||
Qy    728 ATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTC 787

Db    478 TCTCTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATCTA 537
          |||||||||||||||||||||||||| |||||||| |||||||| || ||||| || ||
Qy    788 GCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTA 847

Db    538 CTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTGGAAGATTCCCCATACGAGCC 597
          |||||||||||||||||||||||||||||||| || ||||| |||||||||| || ||
Qy    848 CTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGAACC 907

Db    598 GGTTAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAG-TG------- 649
          ||||||||||  |||||||||||||||||  ||||| ||||||||| ||
Qy    908 AGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTTTCA 967

Db    650 G--A----A-CACATTTCCAAAGGGAACAACTGCCTCGATGCAGCCAAGGCCTGCAACCT 702
          |  |    | ||||||||||||||||||||| |||||||| ||||||||||||||||
Qy    968 GCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCT 1027

Db    703 GGATGACACCTGCAAGAAGTACAGGTCCGCCTACATCACCCCCTGTACCACCAGCATGTC 762
          || ||||  |||||||||||||||||| || ||||||||| || ||||||||  ||||
Qy   1028 CGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTC 1087

Db    763 CAATGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTCAGGCAGTTCTTCGACAAAGT 822
          ||| || |||||||||||||||||||||||| |||||| |||||||||| ||||| ||
Qy   1088 CAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGT 1147

Db    823 TCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACGTCGCCTGCACCGA 882
          || |||||||||||||||| |||||||||||||||||||||||| |||||||||| ||
Qy   1148 CCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGA 1207

Db    883 GAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTG 942
          | || |||||||||| ||||| |||||||||||||||||||  | |||| |||||||||
Qy   1208 GCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTG 1267

Db    943 CCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATTT 1002
          |||||  |||| |||||||||||||| ||||||||||||||||||||||||| |||||
Qy   1268 TTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTT 1327

Db   1003 TTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGC 1062
          |||||||||||||||||||||||||||||||||||||||| ||||||  |||||  ||||||||
Qy   1328 TTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGC 1387

Db   1063 AGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAGA 1122
          ||||||||||| |||||||||| || |||||||||||||||| || ||||||||||
Qy   1388 TGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGA 1447

Db   1123 CTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGGA 1182
          |||||| |||||||||||||| || ||||| || ||||||||||||||| || ||||| ||
Qy   1448 CTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGA 1507

Db   1183 AGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTCA 1242
          ||| ||| |||| ||| ||||||| || |||||||||| |||| ||||||||||||||
Qy   1508 AGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCA 1567

Db   1243 AGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAACCAGCCCCCCCAGTCCAGACCAC 1302
          |||||||||||||||||| |||||||| ||||||| ||||||  |||||| ||||||||
Qy   1568 AGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCAC 1627

Db   1303 CACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCTC 1362
          |||||||| || |||||||| |||| | |||||||||| || ||||||||||| ||
Qy   1628 CACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTC 1687

Db   1363 TGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGCACAGAAGCT 1422
          ||||||||| |||||||| || ||||| |||||||||| ||||| ||||||||||||||
Qy   1688 TGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCT 1747

Db   1423 GAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTACGGAAAGGA 1482
          |||||||||||| ||||||| |||||| |||||| ||||  ||  | |||| | ||| ||
Qy   1748 GAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGA 1807

Db   1483 TGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTG 1542
          |||||||   |||| ||||||||||||||||||||||||||||||||||||| |||||
Qy   1808 AGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTG 1864

Db   1543 CGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTTGTCTGTATC 1602
          |||||||| |||||| ||| ||||| | ||||||| |  |||| || ||| || |
Qy   1865 TGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAAC 1924

Db   1603 A 1603
          |
Qy   1925 A 1925


RESULT    9
LOCUS       AB000800     1415 bp    mRNA            ROD       04-FEB-1997
DEFINITION  Mouse mRNA for GDNF receptor alpha, complete cds.
ACCESSION   AB000800
NID         g1816441
KEYWORDS    GDNF receptor alpha.
SOURCE      Mus musculus Dorsal root ganglion Schwann cell cell_line:IMS32 cDNA
            to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 1415)
  AUTHORS   Watabe,K.
  TITLE     Regulation of GDNF receptor alpha mRNA in Schwann cell cultures
```

```
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 1415)
  AUTHORS   Watabe,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1997) to the DDBJ/EMBL/GenBank databases.
            Kazuhiko Watabe, Tokyo Metropolitan Institute for Neuroscience,
            Neuropathology; Musashidai 2-6, Fuchu, Tokyo 183, Japan
            (E-mail:kazwtb@tmin.ac.jp, Tel:0423-25-3881, Fax:0423-21-8678)
FEATURES             Location/Qualifiers
     source          1. .1415
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="IMS32"
                     /cell_type="Schwann cell"
                     /tissue_type="Dorsal root ganglion"
     primer_bind     1. .20
     CDS             2. .1408
                     /standard_name="glial cell line-derived neurotrophic
                     factor receptor alpha"
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:d1019934"
                     /db_xref="PID:g1816442"
                     /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     VACTERRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTAMTTTAFRIKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPV
                     MVFTALAALLSVSLAETS"
     primer_bind     complement(1396. .1415)
BASE COUNT      351 a    402 c    358 g    304 t
ORIGIN

  Query Match           42.0%;  Score 1061;  DB 23;  Length 1415;
  Best Local Similarity  88.7%;  Pred. No. 0.00e+00;
  Matches  1233;  Conservative    0;  Mismatches 154;  Indels   3;  Gaps   1;

Db      1 CATGTTCCTAGCCACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTGATGTCGGCCGA 60
          ||||||||| || || |||||||||| |||||| ||| |||| |||||  ||||||||||
Qy    539 CATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGA 598

Db     61 GGTGAGTGGTGGGGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGCCTGAAGGAACA 120
           |||||  || || |||||||||| || |||||||||||||||||||||||||||||  ||
       599 AGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCA 658

       121 GAGCTGCAGCACCAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAGGAAACCAACTT 180
           |||||||||||||||||||| || ||||||||||||||||||||| ||||||||
Qy     659 GAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTT 718

Db     181 CAGCCTGACATCCGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATGGAGGCCTTGAA 240
           |||||||  |||||||||| |||||||||||||||||||||||||| |||||||| ||||
Qy     719 CAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAA 778

Db     241 GCAGAAGTCTCTCTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCT 300
           |||||||| |||||||||||||||||||||||| |||||||| |||||||| || ||
Qy     779 GCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCT 838

Db     301 GCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTGGAAGATTCCCC 360
           ||| || |||||||||||||||||||||||||||||||||| || ||||| ||||||||
Qy     839 GCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCC 898

Db     361 ATACGAGCCGGTTAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGA 420
           ||| || || ||||||||||  |||||||||||||||||  ||||| ||||||||||
Qy     899 ATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGA 958

Db     421 TGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTCGATGCAGCCAAGGC 480
           |||||| |||||||||| |||||| |||||||||||||||||| |||||||| |||||
Qy     959 TGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGC 1018

Db     481 CTGCAACCTGGATGACACCTGCAAGAAGTACAGGTCCGCCTACATCACCCCCTGTACCAC 540
```

```
           ||||||||| || |||| |||||||||||||||| || |||||||||| || |||||
Qy    1019 CTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCAC 1078

Db     541 CAGCATGTCCAATGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTCAGGCAGTTCTT 600
           |||| ||||||| || |||||||||||||||||||||||||| |||||| ||||||||||
Qy    1079 CAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTT 1138

Db     601 CGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACGTCGC 660
           ||||| || || |||||||||||||||| |||||||||||||||||||||||||| ||||
Qy    1139 TGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGC 1198

Db     661 CTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAGAACGAGAGAG 720
           |||||| ||| || |||||||||| ||||| |||||||||||||||||||  | ||||
Qy    1199 CTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAA 1258

Db     721 GCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCT 780
           |||||||||  ||||| |||| |||||||||||||| ||||||||||||||||||||||
Qy    1259 GCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCT 1318

Db     781 TGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGA 840
           ||| ||||||||||||||||||||||||||||||||||||||||||||| |||||| ||||
Qy    1319 TGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGA 1378

Db     841 GAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAA 900
           |||||||| |||||||||| |||||||||| || |||||||||||||||| || ||
Qy    1379 AAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAA 1438

Db     901 CTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAA 960
           |||||||||||||| |||||||||||| || ||||| || |||||||||||||||| ||
Qy    1439 CTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAA 1498

Db     961 TGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAA 1020
           ||||| ||||| ||| |||| ||| ||||||| || |||||||||| ||||| |||||
Qy    1499 CGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAA 1558

Db    1021 TGCAATTCAAGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCAGT 1080
           |||||||||||||||||||||||||||| |||||||| ||||||||||||||  ||||||
Qy    1559 TGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGT 1618

Db    1081 CCAGACCACCACTGCCATGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCC 1140
           |||||||||||||||  || |||||||| |||| | |||||||||| || |||||
Qy    1619 ACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCC 1678

Db    1141 AGCAGGCTCTGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGC 1200
           |||||| |||||||||| |||||||| || ||||| |||||||||| ||||| |||||
Qy    1679 AGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGC 1738

Db    1201 ACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTA 1260
           |||||||||||||||||||| ||||||| |||||| |||||| |||  ||  | ||||
Qy    1739 ACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTA 1798

Db    1261 CGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1320
           | ||| || |||||||   |||| |||||||||||||||||||||||||||||||||||
Qy    1799 TGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1855

Db    1321 TCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTT 1380
           ||| ||||| ||||||||| |||||| |||  ||||| | |||||||| |  |||| ||
Qy    1856 TCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATT 1915

Db    1381 GTCTGTATCA 1390
           ||| || ||
Qy    1916 ATCTTTAACA 1925


RESULT   10
LOCUS       RN2072       1392 bp    RNA             ROD       24-OCT-1997
DEFINITION  Rattus norvegicus mRNA for GDNFR-alpha/TrnRl-delta protein,
            complete cds.
ACCESSION   AJ002072
NID         g2564301
KEYWORDS    GDNFR-alpha/TrnRl-delta protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
```

```
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Zhong,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Zhong J., Molecular Neurobiochemistry
            NC7/172, Ruhr University Bochum, Universitaetsstr. 150, D-44780
            Bochum, GERMANY
REFERENCE   2  (bases 1 to 1392)
  AUTHORS   Zhong,J., Annies,M. and Heumann,R.
  TITLE     Rattus norvegicus GDNFR-alpha/TrnRl-delta mRNA, complete cds
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Rattus norvegicus"
                     /strain="Hannover"
                     /db_xref="taxon:10116"
     CDS             1. .1392
                     /codon_start=1
                     /product="GDNFR-alpha/TrnRl-delta protein"
                     /db_xref="PID:e1154274"
                     /db_xref="PID:g2564302"
                     /translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISVEHISKGNNCLDAAKACNLDD
                     TCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                     RRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKEN
                     YADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTA
                     LAALLSVSLAETS"
BASE COUNT      338 a    406 c    356 g    292 t
ORIGIN

  Query Match           38.4%;  Score 970;  DB 23;  Length 1392;
  Best Local Similarity  88.2%;  Pred. No. 0.00e+00;
  Matches  1223;  Conservative    0;  Mismatches 145;  Indels  18;  Gaps   6;

Db      1 ATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAG 60
          |||||||| || || |||||||||||||||| ||| |||| |||||  |||| |||||
Qy    540 ATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAA 599

Db     61 GTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAG 120
          ||||| || |||||||| |||| || |||||||||| |||||||||||||||| |||
Qy    600 GTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAG 659

Db    121 AGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTC 180
          ||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
          660 AGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTC 719

          181 AGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAG 240
          |||||| |||||||||| |||||||||||||||||||| |||||||||||||||| |||||
Qy    720 AGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAG 779

Db    241 CAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTG 300
          |||||||| || |||||||||||||||||||||||| |||||||| |||||||| || |||
Qy    780 CAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTG 839

Db    301 CGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCG 360
          || || |||||||||||||||||||||||||||||||||| || ||||| ||||||||
Qy    840 CGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCA 899

Db    361 TATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAG-T 419
          ||||| || |||||||||| |||||||||||||||||||  ||||| |||||||||| |
Qy    900 TATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGAT 959

Db    420 G-------G--A----A-CACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCC 465
          |       |  |    | |||||| |||||||||||||||||||| ||||| ||||||
Qy    960 GTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCC 1019

Db    466 TGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACC 525
          |||||||| |||||||  || |||||||||||||||| |||||||||| |||||||||
Qy   1020 TGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACC 1079

Db    526 AGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTC 585
          ||| |||||||||| |||||||||||||| |||||||||||||||||| |||||||||
Qy   1080 AGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTT 1139

Db    586 GACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCC 645
          |||||||| |||||||||||||||||||| ||||||||||||||||||||||||||||||
Qy   1140 GACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCC 1199

Db    646 TGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGG 705
          ||||| |||||| |||||||||| |||| || ||||||||||||||||  | |||| |
Qy   1200 TGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAG 1259

Db    706 CCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTT 765
          |||||||| ||| | |||| |||||||||||||| ||||||||||||||||||||||||
Qy   1260 CCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTT 1319

Db    766 GCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAG 825
          || ||||||||||||||||||||||||||||||||||||||||||| |||||| |||||
Qy   1320 GCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAA 1379

Db    826 AACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAAC 885
          |||||||| |||||||||| |||||||||| || |||||||||||||||||| ||||||
Qy   1380 AACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAAC 1439

Db    886 TACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAAT 945
          ||| |||||||||| |||||||| ||||| |||||||||||||||||||||| || ||
Qy   1440 TACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAAC 1499

Db    946 GACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAAT 1005
          ||||| ||||| |||||||||||| ||||||| || |||||||||| ||||| ||||||
Qy   1500 GACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAAT 1559

Db   1006 GCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTC 1065
          |||||||||||||||||||||||||||| ||||||||| |||||||||||||   |||||
Qy   1560 GCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTA 1619

Db   1066 CAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCA 1125
          |||||||||||||||| || |||||||| ||||||| |||||||||| |||||||||
Qy   1620 CAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCA 1679

Db   1126 GCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCT 1185
          |||||||||||||||| || ||||| || ||||| ||||| ||||| ||||| |||||
Qy   1680 GCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCA 1739

Db   1186 CAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTC 1245
          |||||||||||||||||||||||||||| |  |||||||||| ||||  || || |||
Qy   1740 CAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTAT 1799

Db   1246 GGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCT 1305
          | ||| || |||||||   |||| ||||||||||||||||||||||||||||||||||||
Qy   1800 GAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCT 1856

Db   1306 CCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTA 1365
          || |||||  |||||||| |||||| ||| ||| || | ||||| ||  |  |||| |||
Qy   1857 CCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTA 1916

Db   1366 TCTGTA 1371
          ||| ||
Qy   1917 TCTTTA 1922


RESULT   11
LOCUS       GGU90541     3037 bp    mRNA            VRT       25-JUN-1997
DEFINITION  Gallus gallus GDNF receptor alpha (GDNFRalpha) mRNA, complete cds.
ACCESSION   U90541
NID         g2213802
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Gallus.
```

```
REFERENCE   1  (bases 1 to 3037)
  AUTHORS   Buj-Bello,A., Adu,J., Pinon,L.G., Horton,A., Thompson,J.,
            Rosenthal,A., Chinchetru,M., Buchman,V.L. and Davies,A.M.
  TITLE     Neurturin responsiveness requires a GPI-linked receptor and the Ret
            receptor tyrosine kinase
  JOURNAL   Nature 387 (6634), 721-724 (1997)
  MEDLINE   97336104
REFERENCE   2  (bases 1 to 3037)
  AUTHORS   Buchman,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1997) Biol. Med. Sciences, Univ. of St. Andrews,
            Bute Medical Buildings, St. Andrews, Fife KY16 9TS, Scotland
FEATURES             Location/Qualifiers
     source          1. .3037
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            740. .2150
                     /gene="GDNFRalpha"
     CDS             740. .2149
                     /gene="GDNFRalpha"
                     /note="GPI-linked component of GDNF receptor"
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2213803"
                     /translation="MFLALLYLALPLADVLLSAEVSGLPGGDRLDCVKASDQCLKEQS
                     CSTKYRTLRQCVAGKESNFSRATGLEAKDECKSAMEALKQKSLYNCRCKRGMKKEKNC
                     LRIYWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRLAPIVSVEPVLSKGNNCLDAAKAC
                     NLNDTCKRFRSAYITPCTSSTSNEICNKRKCHKALRLFFDKVPPKHSYGMLFCSCRDV
                     ACTERRRQTIVPVCSYEDREKPNCLNLQESCKKNYICRSRLADFFTNCQPESRSVSSC
                     LKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDIDECRKFLNFFQDN
                     TCLKNAIQAFGNGTDVNVWQPILPVQTTTATTTTASRLKNTGSETTNNEIPTHNDSPA
                     CANLQAQKKRKSNESVDTELCLNENAIGKDNTPGVSTSHISSENSFALPTSFYPSTPL
                     ILMTIALSLFLFLSSSVVL"
     polyA_signal    3005. .3010
BASE COUNT      773 a    764 c    722 g    778 t
ORIGIN

  Query Match           24.4%;  Score 615;  DB 17;  Length 3037;
  Best Local Similarity  81.8%;  Pred. No. 0.00e+00;
  Matches   868;  Conservative    0;  Mismatches 181;  Indels  12;  Gaps   6;

Db     814 CGGAGGGGACCGCCTCGACTGCGTGAAAGCCAGCGATCAGTGTCTCAAGGAGCAGAGCTG 873
           ||| || |||||||| || ||||||||||||| |||||||| || ||||||||||||||
Qy     605 CGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTG 664

Db     874 TAGTACTAAGTACAGGACACTGAGGCAGTGTGTAGCCGGCAAAGAGAGCAACTTCAGCCG 933
           || || |||||| | || || |||||||| || || ||||| |||| ||||||||||||
Qy     665 CAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCT 724

Db     934 GGCGACGGGCCTGGAGGCGAAGGATGAATGCAAAAGCGCCATGGAGGCTCTCAAGCAGAA 993
           |||  | |||||||||||| |||||||| |||   ||||||||||||||| || ||||||||
Qy     725 GGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAA 784

Db     994 ATCTCTGTACAACTGCCGCTGCAAGAGGGGCATGAAGAAGGAGAAAAACTGCCTGCGCAT 1053
           || || ||||||||||||||||||| |||| |||||||||||||| ||||||||||||||
Qy     785 GTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCAT 844

Db    1054 TTACTGGAGCATGTACCAGAGCTTACAGGGAAATGACTTGCTTGAGGATTCTCCCTATGA 1113
           |||||||||||||||||||| | ||||||||||  |||| |||||||| || |||||
Qy     845 TTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGA 904

Db    1114 ACCAGTTAACAGCAGGCTATCAGACATATTCAGGCTAGCACCGATTGTATCAG-TGG--- 1169
           |||||||||||||||  | ||||| |||||| || | |  ||  |  |||||| ||
Qy     905 ACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTT 964

Db    1170 --AGCCAGT--A-CT-AT--CAAAGGGGAACAACTGCTTGGATGCAGCAAAAGCTTGTAA 1221
             ||| |||  | |  ||  | || |||||||||||| |||||||||| || || || ||
Qy     965 TCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAA 1024

Db    1222 CCTAAATGATACCTGCAAGAGGTTCAGATCTGCTTACATAACCCCCTGCACCAGCAGCAC 1281
           |||  | || |  ||||||| || ||| || || ||||| ||||| ||||||| ||||
Qy    1025 CCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGT 1084

Db    1282 GTCTAATGAAATCTGTAACAAGCGGAAGTGTCATAAGGCCCTCCGGCTATTTTTTGACAA 1341
           ||| || ||  |||| |||   || ||||| || ||||||||||||  || ||||||||
Qy    1085 GTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAA 1144

Db    1342 AGTTCCCCCAAAGCACAGCTACGGGATGCTCTTCTGCTCCTGTCGAGACGTAGCCTGTAC 1401
           || ||  | |||||||||||||| |||||||||||||||| || ||| | ||||| ||
Qy    1145 GGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCAC 1204

Db    1402 AGAAAGGAGGCGGCAGACTATTGTTCCTGTGTGTTCATATGAGGACAGGGAGAAACCAAA 1461
           |||  ||||||| ||||| || || |||||||| || ||||| || |||||||| || ||
Qy    1205 AGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAA 1264

Db    1462 CTGCCTGAATTTACAAGAATCCTGCAAGAAGAATTACATCTGCAGATCTCGCCTTGCAGA 1521
           |||  ||||||| || || |||||||| |||||||||||||||||||||||||||||| ||
Qy    1265 CTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGA 1324

Db    1522 TTTTTTCACAAACTGCCAGCCTGAGTCACGCTCTGTTAGTAGCTGTCTGAAGGAGAACTA 1581
           |||||| || |||||||||| |||||| | |||| || |||||||| ||||| |||||
Qy    1325 TTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTA 1384

Db    1582 CGCTGACTGCCTCCTCGCTTACTCGGGGCTCATTGGCACAGTGATGACACCGAACTACAT 1641
           |||||||||||||||||| |||||||||| |||||||||| ||||| || ||||||||
Qy    1385 CGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACAT 1444

Db    1642 AGACTCAAGTAGTCTCAGCGTTGCCCCATGGTGTGACTGCAGCAACAGTGGTAATGACAT 1701
           |||||| |||| ||||| || ||||||||||||||||||||||||||||||||||| || ||| |
Qy    1445 AGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCT 1504

Db    1702 AGATGAATGCCGGAAATTTCTGAATTTCTTCCAGGACAACACATGCCTTAAAAATGCAAT 1761
           ||| || |||  |||||| |||||||||| |||||| ||||| ||||||||||||||
Qy    1505 AGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAAT 1564

Db    1762 TCAGGCTTTCGGCAATGGTACTGATGTGAATGTGTGGCAGCCAATATTACCGGTACAGAC 1821
           ||| || || ||||||||  | |||||||  ||||||||||||   || || ||||||||
Qy    1565 TCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGAC 1624

Db    1822 CACTACAGCCACAACTACAACAGCTTCCAGACTTAAAAACA 1862
           ||| || ||||| || || || ||    | |  |||| ||||
Qy    1625 CACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACA 1665


RESULT   12
LOCUS       HSGFRA1G11    832 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) gene, exon 11 and
            complete cds.
ACCESSION   AF038420
NID         g2921543
KEYWORDS    .
SEGMENT     11 of 11
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 832)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
     mRNA            join(AF038410:1. .314,AF038411:111. .396,AF038412:76. .369,
                     AF038413:162. .245,AF038414:131. .145,AF038415:147. .483,
```

```
                AF038416:202. .311,AF038417:132. .266,AF038418:106. .287,
                AF038419:118. .171,80. .832)
                /gene="GFRA1"
                /product="GPI-linked anchor protein"
     gene       order(AF038410:1. .385,AF038411:1. .529,AF038412:1. .497,
                AF038413:1. .363,AF038414:1. .270,AF038415:1. .596,
                AF038416:1. .436,AF038417:1. .397,AF038418:1. .432,
                AF038419:1. .256,1. .832)
                /gene="GFRA1"
     CDS        join(AF038411:357. .396,AF038412:76. .369,AF038413:162. .245,
                AF038414:131. .145,AF038415:147. .483,AF038416:202. .311,
                AF038417:132. .266,AF038418:106. .287,AF038419:118. .171,
                80. .226)
                /gene="GFRA1"
                /note="part of functional ligand for RET receptor tyrosine
                kinase"
                /codon_start=1
                /product="GPI-linked anchor protein"
                /db_xref="PID:g2921545"
                /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKA
                CNLDDICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                IACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSS
                CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKD
                NTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVL
                PPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVL
                VVTALSTLLSLTETS"
     exon       80. .832
                /gene="GFRA1"
                /number=11
     3'UTR      227. .832
                /gene="GFRA1"
BASE COUNT      245 a    160 c    167 g    260 t
ORIGIN

  Query Match           19.4%;  Score 491;  DB 22;  Length 832;
  Best Local Similarity  94.2%;  Pred. No. 2.25e-297;
  Matches   733;  Conservative    0;  Mismatches   5;  Indels  40;  Gaps   2;

Db      80 GGTAATTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCT 139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1791 GGTAATTATGAAAAAGAAGGTCTCGGTGCTTCCAGCCACATAACCACAAAATCAATGGCT 1850

Db     140 GCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACC 199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1851 GCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACC 1910

      200 CTATTATCTTTAACAGAAACATCATGACTGCATTAAAAAAATACAATATGGACATGTAAA 259
           ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
     1911 CTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAA 1970

Db     260 AAGACAAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGG 319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1971 AAGACAAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTTAGG 2030

Db     320 AGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTT-CCTTTTAAGAAA 378
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Qy    2031 AGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTTTTAAGAAA 2090

Db     379 GCTTCTTGTGATCCTTCGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTC 438
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Qy    2091 GCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTC 2150

Db     439 AGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCA 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2151 AGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCA 2210

Db     499 AACTGGGGCTGTGTTTTCGATGATGATGATCATCATGATCATGAT-------------- 543
           |||||||||||||||||||||||||||||| |||||||| |||||
Qy    2211 AACTGGGGCTGTGTTTTCGATGATGATGATNATCATGATNATGATNNNNNNNNNNNNNNN 2270

Db     544 ------------------------GATTTTAACAGTTTTACTTCTGGCCTTTCCTAGCTA 579
                                      ||||||||||||||||||||||||||||||||||
Qy    2271 NNNNNNNNNNNNNNNNNNNNNNNNNGATTTTAACAGTTTTACTTCTGGCCTTTCCTAGCTA 2330

Db     580 GAGAAGGAGTTAATATTTCTAAGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTA 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2331 GAGAAGGAGTTAATATTTCTAAGGTAACTCCCATATCTCCTTTAATGACATTGATTTCTA 2390

Db     640 ATGATATAAATTTCAGCCTACATTGATGCCAAGCTTTTTTGCCACAAAGAAGATTCTTAC 699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2391 ATGATATAAATTTCAGCCTACATTGATGCCAAGCTTTTTTGCCACAAAGAAGATTCTTAC 2450

Db     700 CAAGAGTGGGCTTTGTGGAAACAGCTGGTACTGATGTTCACCTTTATATATGTACTAGCA 759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2451 CAAGAGTGGGCTTTGTGGAAACAGCTGGTACTGATGTTCACCTTTATATATGTACTAGCA 2510

Db     760 TTTTCCACGCTGATGTTTATGTACTGTAAACAGTTCTGCACTCTTGTACAAAAGAAAA 817
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2511 TTTTCCACGCTGATGTTTATGTACTGTAAACAGTTCTGCACTCTTGTACAAAAGAAAA 2568


RESULT   13
LOCUS       AF012811      777 bp    mRNA            ROD       06-AUG-1997
DEFINITION  Mus musculus GDNF receptor alpha mRNA, partial cds.
ACCESSION   AF012811
NID         g2286224
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   Puliti,A. and Seri,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Laboratorio di Genetica Molecolare,
            Istituto Giannina Gaslini, Largo G. Gaslini, 5, Genova 16148, Italy
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     source          1. .430
                     /organism="Mus musculus"
                     /clone="402011"
                     /note="EST deposited under GenBank Accession Number W76716
                     from Soares mouse embryo NbME13.5 14.5 cDNA library; IMAGE
                     Consortium"
     CDS             <1. .680
                     /codon_start=3
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2286225"
                     /translation="LNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLL
                     AYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAF
                     GNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKL
                     KSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLS
                     VSLAETS"
BASE COUNT      202 a    212 c    174 g    189 t
ORIGIN

  Query Match           19.3%;  Score 488;  DB 23;  Length 777;
  Best Local Similarity  84.9%;  Pred. No. 2.39e-295;
  Matches   661;  Conservative    0;  Mismatches 107;  Indels  11;  Gaps   7;

Db       4 TGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATTTTT 63
           ||||| |||| |||||||||||||| |||||||||||||||||||||||| |||||||
Qy    1270 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 1329

Db      64 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAG 123
           |||||||||||||||||||||||||||||||||| ||||| ||||| |||||||| |
Qy    1330 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 1389

Db     124 ACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAGACT 183
           |||||||||| |||||||||| || |||||||||||||||| || ||||||||||||
Qy    1390 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1449
```

```
Db    184 CCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGGAAG 243
          |||| |||||||||||||| || ||||| || |||||||||||||| || ||||| ||||
Qy   1450 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1509

Db    244 ATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTCAAG 303
          | ||| |||| ||| ||||||| || |||||||||||| ||||| ||||||||||||||||
Qy   1510 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1569

Db    304 CCTTTGGCAATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCAGTCCAGACCACCA 363
          |||||||||||||||| |||||||| |||||||||||||  |||||| ||||||||||
Qy   1570 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1629

Db    364 CTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCTCTG 423
          ||||||| || |||||||| ||||| | |||||||||| || |||||||||| ||||
Qy   1630 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1689

Db    424 AGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGCACAGAAGCTGA 483
          ||||||| |||||||| || ||||| |||||||||| ||||| ||||||||||||||||
      1690 AGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGA 1749

Db    484 AATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTACGGAAAGGATG 543
          |||||||||| ||||||| |||||| |||||| |||| || | |||| | ||| || | |
Qy   1750 AATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAG 1809

Db    544 GTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTGCG 603
          ||||||   |||| ||||||||||||||||||||||||||||||||||| ||||| |
Qy   1810 GTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTG 1866

Db    604 GTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTTGTCTGTATCAT 663
          |||||||| |||||| |||  ||||| | ||||||||| |  |||| || ||| || ||
Qy   1867 GTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACAG 1926

Db    664 TGGCAGAA-ACAT-CGT-AGCTGCATCCAGTATGAAAA-GACAAAAGA---AAACCAAGT 716
            ||  | |  | | | |     || || |||| | | |  ||||||   |||||||||
Qy   1927 AAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGACAAAAACCAAGT 1986

Db    717 ATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCTCGGTTGAGAA 775
          |||||  ||||| |||||||||| |||||| |||||||| |  ||||| ||||||||
Qy   1987 TATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTT-AGGAGCTCAGTTGAGAA 2044
```

```
RESULT   14
LOCUS       HSGFRA1G06    596 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) gene, exon 6.
ACCESSION   AF038415
            g2921538
KEYWORDS    .
SEGMENT     6 of 11
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 596)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
     exon            147. .483
                     /gene="GFRA1"
                     /number=6
BASE COUNT      152 a    156 c    163 g    125 t
ORIGIN
```

```
  Query Match           13.4%;  Score 339;  DB 22;  Length 596;
  Best Local Similarity 100.0%;  Pred. No. 3.42e-195;
  Matches   339;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    145 AGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGA 204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    971 AGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGA 1030

Db    205 CGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCAA 264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1031 CGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCAA 1090

Db    265 CGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCC 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1091 CGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCC 1150

Db    325 GGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCG 384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1151 GGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCG 1210

Db    385 GAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTT 444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1211 GAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTT 1270

Db    445 GAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAG 483
          |||||||||||||||||||||||||||||||||||||||
Qy   1271 GAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAG 1309
```

```
RESULT   15
LOCUS       HSGFRA1G01    385 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) gene, exon 1.
ACCESSION   AF038410
NID         g2921533
KEYWORDS    .
SEGMENT     1 of 11
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 385)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
     exon            1. .314
                     /gene="GFRA1"
                     /number=1
BASE COUNT       83 a    111 c     91 g    100 t
ORIGIN
```

```
  Query Match           11.8%;  Score 298;  DB 22;  Length 385;
  Best Local Similarity 100.0%;  Pred. No. 7.33e-168;
  Matches   298;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     18 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGCTTCCAATAACCACTAACATCCCTAA 77
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGCTTCCAATAACCACTAACATCCCTAA 60

Db     78 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGAG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGAG 120

Db    138 CTCTCGAAGATTACCGCATCTATTTTTTTTTTCTTTTTTTTCTTTTCCTAGCGCAGATAA 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 CTCTCGAAGATTACCGCATCTATTTTTTTTTTCTTTTTTTTCTTTTCCTAGCGCAGATAA 180

Db    198 AGTGAGCCCGGAAAGGGAAGGAGGGGGCGGGGACACCATTGCCCTGAAAGAATAAATAAG 257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 AGTGAGCCCGGAAAGGGAAGGAGGGGGCGGGGACACCATTGCCCTGAAAGAATAAATAAG 240

Db    258 TAAATAAACAAACTGGCTCCTCGCCGCAGCTGGACGCGGTCGGTTGAGTCCAGGTTGG 315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 TAAATAAACAAACTGGCTCCTCGCCGCAGCTGGACGCGGTCGGTTGAGTCCAGGTTGG 298

rch completed: Wed Jun 24 19:01:51 1998
time : 5592 secs.
```

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jun 24 20:29:12 1998; MasPar time 376.73 Seconds
868.165 Million cell updates/sec
ular output not generated.

Title: >US-08-866-354-1
Description: (1-2568) from US08866354.seq
Perfect Score: 2525
N.A. Sequence: 1 AATCTGGCCTCGGAACACGC.....CACTCTTGTACAAAAGAAAA 2568
Comp: TTAGACCGGAGCCTTGTGCG.....GTGAGAACATGTTTTCTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 176886 seqs, 63680241 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37

Statistics: Mean 9.999; Variance 6.681; scale 1.497

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	1137	45.0	2378	37	T84975	Rat glial cell derive	0.00e+00
	2	612	24.2	840	37	T84977	Mouse GDNFR alpha clo	0.00e+00
	3	226	9.0	418	37	T84976	Mouse GDNFR alpha clo	1.22e-117
	4	131	5.2	453	37	T84979	Human EST-derived seq	3.49e-59
	5	124	4.9	351	37	T84978	Human EST-derived seq	5.78e-55
	6	96	3.8	385	15	Q85526	DNA probe 32 detects	2.63e-38
	7	90	3.6	201	37	T84981	Human EST-derived seq	8.55e-35
	8	90	3.6	201	37	T84980	Human EST-derived seq	8.55e-35
	9	45	1.8	204	1	N81164	Base substituted E.co	1.07e-09
	10	44	1.7	91	9	Q51746	Oligonucleotide probe	3.52e-09
c	11	42	1.7	91	9	Q51746	Oligonucleotide probe	3.72e-08
c	12	43	1.7	204	1	N81164	Base substituted E.co	1.15e-08
	13	39	1.5	114	12	Q70469	Generic DNA sequence	1.20e-06
	14	38	1.5	114	12	Q70467	Generic DNA sequence	3.76e-06
c	15	39	1.5	172	32	T76363	Human interleukin 8 a	1.20e-06
	16	36	1.4	114	12	Q70468	Generic DNA sequence	3.59e-05
	17	36	1.4	114	12	Q70465	Generic DNA sequence	3.59e-05
	18	36	1.4	114	12	Q70466	Generic DNA sequence	3.59e-05
c	19	36	1.4	114	12	Q70467	Generic DNA sequence	3.59e-05
c	20	36	1.4	114	12	Q70468	Generic DNA sequence	3.59e-05
c	21	35	1.4	114	12	Q70465	Generic DNA sequence	1.09e-04
c	22	35	1.4	114	12	Q70469	Generic DNA sequence	1.09e-04
	23	34	1.3	114	12	Q70470	Generic DNA sequence	3.29e-04
	24	34	1.3	114	12	Q70471	Generic DNA sequence	3.29e-04
	25	33	1.3	114	12	Q70472	Generic DNA sequence	9.79e-04
	26	32	1.3	114	12	Q70473	Generic DNA sequence	2.88e-03
c	27	34	1.3	114	12	Q70466	Generic DNA sequence	3.29e-04
c	28	33	1.3	114	12	Q70470	Generic DNA sequence	9.79e-04
c	29	32	1.3	168	31	T76270	Human MDNCF antisense	2.88e-03
	30	32	1.3	172	32	T76363	Human interleukin 8 a	2.88e-03
	31	32	1.3	178	31	T76405	Human endothelin-1 an	2.88e-03
	32	32	1.3	190	31	T76452	Chymase antisense oli	2.88e-03
c	33	30	1.2	114	12	Q70472	Generic DNA sequence	2.39e-02
c	34	30	1.2	178	31	T76405	Human endothelin-1 an	2.39e-02
	35	30	1.2	250	19	T22691	Human gene signature	2.39e-02
	36	30	1.2	8920	11	Q62924	Carbamoyl-phosphate-s	2.39e-02
c	37	29	1.1	114	12	Q70473	Generic DNA sequence	6.75e-02
c	38	27	1.1	128	31	T76233	Human IL6 antisense o	5.13e-01
	39	27	1.1	168	31	T76270	Human MDNCF antisense	5.13e-01
c	40	28	1.1	416	8	Q59427	Human brain Expressed	1.88e-01
	41	29	1.1	598	1	N90919	Type A insertion gene	6.75e-02
	42	29	1.1	1021	16	T04613	5' flanking region of	6.75e-02
	43	29	1.1	2454	15	T01567	Hamster cDNA probe sp	6.75e-02
c	44	27	1.1	2504	30	T62359	Schizosaccharomyces p	5.13e-01
	45	29	1.1	4641	1	N81538	Sequence of a gene fr	6.75e-02

ALIGNMENTS

RESULT 1
ID T84975 standard; cDNA; 2378 BP.
AC T84975;
DT 27-APR-1998 (first entry)
DE Rat glial cell derived neurotrophic factor receptor alpha cDNA.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; rat; kidney disease; glomerulonephritis; therapy; ds.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 117..1523
FT /*tag= a
FT sig_peptide 117..188
FT /*tag= b
FT mat_peptide 189..1520
FT /*tag= c
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
DR P-PSDB; W27327.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Claim 23; Page 74-77; 100pp; English.
CC This cDNA codes for full-length rat glial cell derived neurotrophic
CC factor receptor alpha (GDNFR alpha) (see W27327), a novel
CC GPI-linked protein that is a ligand-binding component of the
CC receptor system for GDNF. It was isolated by expression cloning.
CC A cDNA library obtained from ventral midbrain tissue of E14 rat
CC embryos was generated in a cytomegalovirus-based vector. cDNA
CC clones were transfected into COS 7 cells and expression of
CC putative GDNF receptors was detected by binding of iodinated GDNF.
CC A single positive pool was obtained, from which the cDNA clone was
CC isolated. An expression vector containing the cDNA can be used to

```
CC   produce GDNFR alpha in transformed host cells.  The invention
CC   relates to novel uses of GDNF and its receptor.  In particular, it
CC   relates to native rat GDNFR alpha (see W27327), its variants and
CC   soluble derivatives (extracellular domain), chimeric GDNFR alpha
CC   and antibodies which bind to the GDNFR alpha, including agonist
CC   and neutralising antibodies, as well as various uses for these
CC   molecules.  It also relates to assay systems for detecting ligands
CC   to GDNFR alpha, systems for studying the physiological role of
CC   GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC   methods for identifying molecules homologous to GDNFR alpha, and
CC   therapeutic techniques (claimed) for the treatment of GDNF-related
CC   and GDNFR alpha-related conditions, particularly kidney disease
CC   associated with glomerulonephritis and enteric nervous system
CC   related disorders.  Transgenic and knockout animals are also
CC   claimed.
SQ   Sequence   2378 BP;    617 A;    624 C;    598 G;    539 T;

  Query Match           45.0%;  Score 1137;  DB 37;  Length 2378;
  Best Local Similarity  84.2%;  Pred. No. 0.00e+00;
  Matches  1481;  Conservative    0;  Mismatches 263;  Indels  14;  Gaps  10;

Db     81 ttttgggggggcggggaccagctgcgcggcggcaccatgttcctagccactctgtacttc 140
Qy    504 TTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTC 563

Db    141 gcgctgccactcctggatttgctgatgtccgccgaggtgagtggtggagaccgtctggac 200
Qy    564 GCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGAT 623

Db    201 tgtgtgaaagccagcgatcagtgcctgaaggaacagagctgcagcaccaagtaccgcaca 260
Qy    624 TGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACG 683

Db    261 ctaaggcagtgcgtggcgggcaaggaaaccaacttcagcctgacatccggccttgaggcc 320
Qy    684 CTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCC 743

Db    321 aaggatgagtgccgtagcgccatggaggccttgaagcagaagtctctgtacaactgccgc 380
Qy    744 AAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGC 803

Db    381 tgcaagcggggcatgaagaaagagaagaattgtctgcgtatctactggagcatgtaccag 440
Qy    804 TGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTACTGGAGCATGTACCAG 863

Db    441 agcctgcagggaaatgacctcctggaagattccccgtatgagccggttaacagcaggttg 500
Qy    864 AGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGAACCAGTTAACAGCAGATTG 923

Db    501 tcagatatattccgggcagtcccgttcatatcagatgttttccagcaagtggaacacatt 560
Qy    924 TCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTTTCAGCAAGTGGAGCACATT 983

Db    561 tccaaagggaacaactgcctggacgcagccaaggcctgcaacctggacgacacctgtaag 620
Qy    984 CCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAG 1043

Db    621 aagtacaggtcggcctacatcacccccctgcaccaccagcatgtccaacgaggtctgcaac 680
Qy   1044 AAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCAACGATGTCTGCAAC 1103

Db    681 cgccgtaagtgccacaaggccctcaggcagttcttcgacaaggttccggccaagcacagc 740
Qy   1104 CGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGC 1163

Db    741 tacgggatgctcttctgctcctgccgggacatcgcctgcaccgagcggcggcgacagact 800
Qy   1164 TACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACC 1223

Db    801 atcgtccccgtgtgctcctatgaagaacgagagaggcccaactgcctgagtctgcaagac 860
Qy   1224 ATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGAC 1283

Db    861 tcctgcaagaccaattatatctgcagatctcgccttgcagatttttttaccaactgccag 920
Qy   1284 TCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAG 1343

Db    921 ccagagtcaaggtctgtcagcaactgtcttaaggagaactacgcagactgcctcctggcc 980
Qy   1344 CCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCC 1403

Db    981 tactcgggactgattggcacagtcatgactcccaactacgtagactccagcagcctcagc 1040
Qy   1404 TACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACTCCAGTAGCCTCAGT 1463

Db   1041 gtggcaccatggtgtgactgcagcaacagcggcaatgacctggaagactgcttgaaattt 1100
Qy   1464 GTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTT 1523

Db   1101 ctgaatttttttaaggacaatacttgtctcaaaaatgcaattcaagcctttggcaatggc 1160
Qy   1524 TTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGC 1583

Db   1161 tcagatgtgaccatgtggcagccagcccctccagtccagaccaccactgccaccactacc 1220
Qy   1584 TCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCACTGCCACTACCACC 1643

Db   1221 actgccttccgggtcaagaacaagcctctggggccagcagggtctgagaatgagatcccc 1280
Qy   1644 ACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCC 1703

Db   1281 acacacgttttaccaccctgtgcgaatttgcaggctcagaagctgaaatccaatgtgtcg 1340
Qy   1704 ACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCG 1763

Db   1341 ggtagcacacacctctgtctttctgatagtgatttcggaaaggatggtctcgctggtgcc 1400
Qy   1764 GGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAGGTCTCG---GTGCT 1820

Db   1401 tccagccacataaccacaaaatcaatggctgctcctcccagctgcagtctgagctcactg 1460
Qy   1821 TCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTG 1880

Db   1461 ccggtgctgatgctcaccgcccttgctgccctgttatctgtatcgttggcagaaacgtcg 1520
Qy   1881 CTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACAGAAACATCATAGCTG 1940

Db   1521 tagctgcatccgggaaaacagtatgaaaagacaaaagagaaccaagtattctgtc-cctg 1579
Qy   1941 CATTAAAAAAATACAATATGGACATGTAAAAAGACAAA-AACCAAGTTATCTGTTTCCTG 1999

Db   1580 tcctcttgtatatctgaaaatccagttttaaaagctccgttgagaagcagtttcacccaa 1639
Qy   2000 TTCTCTTGTATAGCTGAAATTCCAGTTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAA 2058

Db   1640 ctggaac-tctttccttgt---ttttaagaaagcttgt-g-gc-cctcaggggcttctgt 1692
Qy   2059 CTGGAACATTTTTTTTTTTNCCTTTTAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGT 2118

Db   1693 tgaagaactgctacagggctaattccaaacccataaggctctggggcgtggtgcggctta 1752
Qy   2119 GAAAAACCTGATGCAGTGCTCCATCCAAACTCAGAAGGCTTTGGGATATGCTGTATTTTA 2178

Db   1753 aggggaccatttgcaccatgtaaagcaagctgggcttatcatgtgtttgatggtgaggat 1812
Qy   2179 AAGGGACAGTTTGTAACTTGGGCTGTAAAGCAAACTGGGGCTGTGTTTTC-GATGATGAT 2237

Db   1813 ggtagtggtgatgatgat 1830
Qy   2238 GATNATCATGATNATGAT 2255


RESULT    2
ID   T84977 standard; cDNA; 840 BP.
AC   T84977;
DT   27-APR-1998  (first entry)
```

DE Mouse GDNFR alpha clone 26 3' end.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; mouse; kidney disease; glomerulonephritis; therapy; ss.
OS Mus musculus.
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Example 1; Page 60-61; 100pp; English.
CC This cDNA sequence comprises the 3' end of mouse full-length
CC glial cell derived neurotrophic factor receptor alpha (GDNFR alpha)
CC (see W27327) clone 26; the 5' end of the clone is given in
CC T84978. Clone 26 was isolated from a mouse cDNA library using
rat GDNFR alpha cDNA (see T84975) as probe. The invention relates
to novel uses of GDNF and its receptor. In particular, it relates
to native rat GDNFR alpha (see W27327), its variants and soluble
CC derivatives (extracellular domain), chimeric GDNFR alpha and
CC antibodies which bind to the GDNFR alpha, including agonist and
CC neutralising antibodies, as well as various uses for these
CC molecules. It also relates to assay systems for detecting ligands
CC to GDNFR alpha, systems for studying the physiological role of
CC GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC methods for identifying molecules homologous to GDNFR alpha, and
CC therapeutic techniques (claimed) for the treatment of GDNF-related
CC and GDNFR alpha-related conditions, particularly kidney disease
CC associated with glomerulonephritis and enteric nervous system
CC related disorders. Transgenic and knockout animals are also
CC claimed.
SQ Sequence 840 BP; 205 A; 245 C; 200 G; 190 T;

Query Match 24.2%; Score 612; DB 37; Length 840;
Best Local Similarity 88.0%; Pred. No. 0.00e+00;
Matches 726; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

```
Db      1 cgccgcaagtgccacaaagccctcaggcagttcttcgacaaagttccagccaagcacagc 60
          |||||||||||||||||| |||||| |||||||||| ||||| || || ||||||||||||
Qy   1104 CGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGC 1163

Db     61 tacgggatgctctttctgctcctgccgggacgtcgcctgcaccgagaggcggcgacagact 120
          |||||  |||||||||||||||||||||||||| ||||||||||  ||| || ||||||||||
Qy   1164 TACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACC 1223

Db    121 atcgtccctgtgtgctcctatgaagaacgagagaggcccaactgcctgaatctgcaagac 180
          |||||  |||||||||||||||||||  | ||||  |||||||||  |||||  ||||  |||
Qy   1224 ATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGAC 1283

Db    181 tcctgcaagacaaattacatctgcagatctcgccttgcagatttttttaccaactgccag 240
          ||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||
Qy   1284 TCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAG 1343

Db    241 ccagagtcaaggtctgtcagcaactgtcttaaggagaactacgcagactgcctcctggcc 300
          |||||||||||||||||||||| |||||| ||||| |||||||| |||||||||| |||
Qy   1344 CCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCC 1403

Db    301 tactcgggactgattggcacagtcatgactcctaactacatagactccagcagcctcagt 360
          ||||||||  || |||||||||||||||||  || |||||||||||||||||  |||||||||
Qy   1404 TACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACTCCAGTAGCCTCAGT 1463

Db    361 gtggcgccgtggtgcgattgcagcaacagtggcaatgacctggaagattgcctgaagttt 420
          |||||  ||  |||||  ||  |||||||||||||||  ||  |||||  |||||  |||  ||||  |||
Qy   1464 GTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTT 1523

Db    421 ctgaatttttttaaggacaatacgtgtctcaaaaatgcaattcaagcctttggcaatggc 480
          |||||||  ||  ||||||||||||  |||||  ||||||||||||||||||||||||||||
Qy   1524 TTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGC 1583

Db    481 tcggatgtgaccatgtggcagccagccccccccagtccagaccaccactgccacgactacc 540
          ||  |||||||||| ||||||||||||||||  ||||||  |||||||||||||||||  ||  |||
Qy   1584 TCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCACTGCCACTACCACC 1643

Db    541 actgccttccggatcaagaacaagcctctagggccagcaggctctgagaatgagattccc 600
          ||||||  |||||  |  ||||||||||  ||  ||||||||||  ||||||||||  ||||||
Qy   1644 ACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCC 1703

Db    601 acacacgttttaccaccgtgtgctaatttgcaggcacagaagctgaaatccaatgtatcg 660
          ||  ||  |||||  ||||||||||  |||||  |||||||||||||||||||||||||  |||
Qy   1704 ACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCG 1763

Db    661 ggcagtacacatctctgtctttctgataatgattacggaaaggatggtctcgctggtgcc 720
          ||||  ||||||  ||||||  ||||  ||  |  ||||  |  |||  ||  |||||||  ||||
Qy   1764 GGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAGGTCTCG---GTGCT 1820

Db    721 tccagccacataaccacaaaatcaatggctgctcctcccagctgcggtctgagctcactg 780
          ||||||||||||||||||||||||||||||||||||||||  |||||  |||||||||  |||||
Qy   1821 TCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTG 1880

Db    781 ccggtgatggtgttcaccgctctggctgccctgttgtctgtatca 825
          |  |||  |||||  |  |||||||||  |  ||||  ||  |||  ||  ||
Qy   1881 CTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACA 1925
```

RESULT 3
ID T84976 standard; cDNA; 418 BP.
AC T84976;
DT 27-APR-1998 (first entry)
DE Mouse GDNFR alpha clone 26 5' end.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; mouse; kidney disease; glomerulonephritis; therapy; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 287..418
FT /*tag= a
FT /note= "5' end of GDNFR alpha coding region"
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Example 1; Page 60; 100pp; English.
CC This cDNA sequence comprises the 5' end of mouse full-length
CC glial cell derived neurotrophic factor receptor alpha (GDNFR alpha)
CC (see W27327) clone 26; the 3' end of the clone is given in
CC T84977. Clone 26 was isolated from a mouse cDNA library using
CC rat GDNFR alpha cDNA (see T84975) as probe. The invention relates
CC to novel uses of GDNF and its receptor. In particular, it relates
CC to native rat GDNFR alpha (see W27327), its variants and soluble
CC derivatives (extracellular domain), chimeric GDNFR alpha and
CC antibodies which bind to the GDNFR alpha, including agonist and
CC neutralising antibodies, as well as various uses for these
CC molecules. It also relates to assay systems for detecting ligands
CC to GDNFR alpha, systems for studying the physiological role of
CC GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC methods for identifying molecules homologous to GDNFR alpha, and
CC therapeutic techniques (claimed) for the treatment of GDNF-related
CC and GDNFR alpha-related conditions, particularly kidney disease
CC associated with glomerulonephritis and enteric nervous system
CC related disorders. Transgenic and knockout animals are also
CC claimed.
SQ Sequence 418 BP; 70 A; 119 C; 150 G; 79 T;

Query Match 9.0%; Score 226; DB 37; Length 418;
Best Local Similarity 83.3%; Pred. No. 1.22e-117;
Matches 320; Conservative 0; Mismatches 58; Indels 6; Gaps 4;

```
Db      38 caggttgggtcggaactgaacccctaaaagcgggtccgcctcccgccctcgcgcccgctc 97
           ||||||||||||||| |||||||||||||||||  ||||||||||||||||    | | |
Qy     291 CAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGCCATC-C-C 348

Db      98 ggagctgagtccctggcggcggtgggcggcagagcaacggggagtctgctctcaccctgg 157
           |||||||||||| | |||||||||||| | || |     ||  ||||| | ||||  | ||||
Qy     349 GGAGCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGG 407

Db     158 atggagcttaactttgagtggccagaggagcgcagtcgcccggggatctctgcacgctga 217
           |||||||| ||||||| | |||||||| ||| |||  | |||||||||  ||||||||||
Qy     408 ATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGA 467

Db     218 gctctctccgcgagatccggtggcggctttggatttt---gggggggcggggaccagctg 274
           |||| ||| || ||| || | |||||||  |||||||   ||||||||||||||||||
Qy     468 GCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCC 527

       275 cgcggtggcaccatgttcctagccactctgtacttcgtgctgccactcctggatttgctg 334
           ||||  |||||||||||||| || || |||||||||| |||||| ||| |||| |||||
       528 CGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTC 587

Db     335 atgtcggccgaggtgagtggtggggaccgcctggactgtgtgaaagccagtgatcagtgc 394
           |||||||||| ||||| || || |||||||||| || ||||||||||||||||||||||
Qy     588 CTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGC 647

Db     395 ctgaaggaacagagctgcagcacc 418
           |||||||| ||||||||||||||||
Qy     648 CTGAAGGAGCAGAGCTGCAGCACC 671
```

RESULT 4
ID T84979 standard; DNA; 453 BP.
AC T84979;
DT 27-APR-1998 (first entry)
DE Human EST-derived sequence yI70a10.r1.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; human; kidney disease; glomerulonephritis; therapy;
KW expressed sequence tag; EST; ss.
OS Homo sapiens.
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
WPI; 97-470819/43.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Example 1; Page 61; 100pp; English.
CC This DNA sequence comprises a human EST-derived sequence designated
CC yI70a10.r1. It can be used to identify glial cell derived
CC neurotrophic factor receptor (GDNFR) sequences, including human
CC variants. Also suitable as probes for GDNFR sequences are another
CC human EST-derived sequence designated ye83h05.r1 (see T84978), and
CC fragments of these EST-derived sequences (see T84980-81) or
CC proteins encoded by them. The invention relates to novel uses of
CC GDNF and its receptor. In particular, it relates to native rat
CC GDNFR alpha (see W27327), its variants and soluble derivatives
CC (extracellular domain), chimeric GDNFR alpha and antibodies which
CC bind to the GDNFR alpha, including agonist and neutralising
CC antiodies, as well as various uses for these molecules. It also
CC relates to assay systems for detecting ligands to GDNFR alpha,
CC systems for studying the physiological role of GDNF, diagnostic
CC techniques for identifying GDNF-related conditions, methods for
CC identifying molecules homologous to GDNFR alpha, and claimed
CC methods for the treatment of GDNF-related and GDNFR alpha-related
CC conditions, particularly kidney disease associated with
CC glomerulonephritis and enteric nervous system related disorders.
CC Transgenic and knockout animals are also claimed.
SQ Sequence 453 BP; 86 A; 152 C; 121 G; 91 T;

Query Match 5.2%; Score 131; DB 37; Length 453;
Best Local Similarity 71.4%; Pred. No. 3.49e-59;
Matches 242; Conservative 0; Mismatches 94; Indels 3; Gaps 3;

```
Db       3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
           ||| | |||||||||| || ||||||||||||| | ||||  |||||||||    |
Qy     993 AACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGG 1052

Db      63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
           ||  | |||||| ||   |||| || | ||   | | ||| |||   ||||||||||||
Qy    1053 TCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCAACCGCCGC 1109

Db     123 aagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
           |||||||||||||||| || |||||||| |||  ||| ||   | || ||| |||| |
Qy    1110 AAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGA 1169

Db     183 atgctcttctgctcctgccaagaccaggcgtgcgctgagcgcgcgggcaaaaccatcctg 242
           |||||||||||||||||  |||   || ||| | |||||   | |  | ||||||  ||
Qy    1170 ATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTG 1229

Db     243 cccagctgctcctatgaggacaaggagaagcccaactgcctggacctgcgtggcgtgtgc 302
           ||    |||||||||| || | |||||||||||||||  || |  |||  | |   |||
Qy    1230 CCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGC 1289

Db     303 cggactgaccacctgtgtcggtcccggctngccgacttt 341
           |||  |  || | ||  | || || || || || |||
Qy    1290 AAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTT 1328
```

RESULT 5
ID T84978 standard; DNA; 351 BP.
AC T84978;
DT 27-APR-1998 (first entry)
DE Human EST-derived sequence ye83h05.r1.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW expressed sequence tag; ss.
OS Homo sapiens.
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Example 1; Page 61; 100pp; English.
CC This DNA sequence comprises a human EST-derived sequence designated
CC ye83h05.r1. It can be used to identify glial cell derived
CC neurotrophic factor receptor (GDNFR) sequences, including human
CC variants. Also suitable as probes for GDNFR sequences are another
CC human EST-derived sequence designated yI70a10.r1 (see T84979), and
CC fragments of these EST-derived sequences (see T84980-81) or
CC proteins encoded by them. The invention relates to novel uses of
CC GDNF and its receptor. In particular, it relates to native rat
CC GDNFR alpha (see W27327), its variants and soluble derivatives
CC (extracellular domain), chimeric GDNFR alpha and antibodies which
CC bind to the GDNFR alpha, including agonist and neutralising
CC antiodies, as well as various uses for these molecules. It also
CC relates to assay systems for detecting ligands to GDNFR alpha,
CC systems for studying the physiological role of GDNF, diagnostic
CC techniques for identifying GDNF-related conditions, methods for
CC identifying molecules homologous to GDNFR alpha, and claimed
CC methods for the treatment of GDNF-related and GDNFR alpha-related
CC conditions, particularly kidney disease associated with
CC glomerulonephritis and enteric nervous system related disorders.
CC Transgenic and knockout animals are also claimed.
SQ Sequence 351 BP; 77 A; 117 C; 96 G; 57 T;

Query Match 4.9%; Score 124; DB 37; Length 351;

```
  Best Local Similarity  72.7%;  Pred. No. 5.78e-55;
  Matches   216;  Conservative    0;  Mismatches  78;  Indels   3;  Gaps   3;

Db      40 ccaagagcaaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaaga 99
           ||||  | ||| | |||||||||| || ||||||||||||||  | ||||  |||||||
Qy     985 CCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGA 1044

Db     100 agctgcgctcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgca 159
           ||    | ||  | |||||| ||   |||| || | ||   | | ||| |||   |||||
Qy    1045 AGTACAGGTCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCA 1101

Db     160 accgccgcaagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtaca 219
           |||||||||||||||||||||||||| || |||||||| |||  ||| ||   | || |||
Qy    1102 ACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACA 1161

Db     220 cctaccgcatgctcttctgctcctgccaagatcaggcgtgcgctgagcncgcggncaaaa 279
           |||| | ||||||||||||||||||  ||    || ||| | ||||    |    | |
Qy    1162 GCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGA 1221

Db     280 ccatcctgcccagctgctcctatgaggacaaggagaagcccaactgcntggacntgc 336
           ||||| ||||    |||||||||| || | ||||||||||||||  || |  |||
      1222 CCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGC 1278


RESULT    6
ID   Q85526 standard; DNA; 385 BP.
AC   Q85526;
DT   10-JAN-1996  (first entry)
DE   DNA probe 32 detects DNA-protein complex in immortal cells.
KW   DNA-protein complex; detection; proliferation; tumour formation;
KW   diagnose; malignancy; biopsy; probe; ss.
OS   Synthetic.
PN   WO9502701-A1.
PD   26-JAN-1995.
PF   13-JUL-1994; E02307.
PR   15-JUL-1993; DE-323727.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Abken H,  Albert W,  Jungfer H,  Abken HJ;
DR   WPI; 95-067344/09.
PT   New DNA-protein complex characteristic of cells with unlimited
PT   proliferation capacity  - and its components and derived antibodies,
PT   useful in diagnosing malignant tumours.
PS   Claim 1; Page 73; 106pp; German.
CC   The DNA shown is found in human or animal cells that have an unlimited
CC   capacity for unlimited cell proliferation or tumour formation. They have
CC   no ability to promote immortalisation of the cells, and are usually
CC   found in a DNA-protein complex in the cell cytoplasm. The DNA is useful
CC   for detection of these complexes, and diagnosis of malignant tumours.
CC   Differentiation between cells with unlimited and only transient
     proliferation is possible. (See also Q85493-54)
     Sequence   385 BP;    121 A;    83 C;    90 G;    91 T;

  Query Match            3.8%;  Score 96;  DB 15;  Length 385;
  Best Local Similarity  92.9%;  Pred. No. 2.63e-38;
  Matches   104;  Conservative    0;  Mismatches   8;  Indels   0;  Gaps   0;

Db       1 gatctcgccttgcagatttttttaccaactgccagccagagtcaaggtctgtcagcaact 60
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||| ||
Qy    1309 GATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCT 1368

Db      61 gtcttaaggagaactacgcagactgcctcctggcctactcgggactgattgg 112
           |||| ||||| |||||||| |||||||||| |||||||||| || |||||
Qy    1369 GTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGG 1420


RESULT    7
ID   T84981 standard; DNA; 201 BP.
AC   T84981;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence yI70a10.r1 fragment.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW   expressed sequence tag; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a fragment of a human EST-derived
CC   sequence designated yI70a10.r1 (see T84979).  It can be used as
CC   a probe to identify glial cell derived neurotrophic factor receptor
CC   (GDNFR) sequences, including human variants.  The invention relates
CC   to novel uses of GDNF and its receptor.  In particular, it relates
CC   to rat GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
CC   Transgenic and knockout animals are also claimed.
SQ   Sequence   201 BP;    41 A;    78 C;    46 G;    36 T;

  Query Match            3.6%;  Score 90;  DB 37;  Length 201;
  Best Local Similarity  76.4%;  Pred. No. 8.55e-35;
  Matches   152;  Conservative    0;  Mismatches  44;  Indels   3;  Gaps   3;

Db       3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
           ||| | |||||||||||| || ||||||||||||||  | ||||  |||||||||    |
Qy     993 AACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGG 1052

Db      63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
           ||  | |||||| ||   |||| || | ||   | | ||| |||   ||||||||||||
Qy    1053 TCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCAACCGCCGC 1109

Db     123 aagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
           |||||||||||||||| || |||||||| |||  ||| ||   | || ||| |||| |
Qy    1110 AAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGA 1169

Db     183 atgctcttctgctcctgcc 201
           ||||||||||||||||||
Qy    1170 ATGCTCTTCTGCTCCTGCC 1188


RESULT    8
ID   T84980 standard; DNA; 201 BP.
AC   T84980;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence ye83h05.r1 fragment.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW   expressed sequence tag; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
```

```
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a fragment of a human EST-derived
CC   sequence designated ye83h05.r1 (see T84978).  It can be used as
CC   a probe to identify glial cell derived neurotrophic factor receptor
CC   (GDNFR) sequences, including human variants.  The invention relates
CC   to novel uses of GDNF and its receptor.  In particular, it relates
CC   to rat GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
     Transgenic and knockout animals are also claimed.
     Sequence   201 BP;    41 A;    78 C;    46 G;    36 T;

  Query Match            3.6%;  Score 90;  DB 37;  Length 201;
  Best Local Similarity  76.4%;  Pred. No. 8.55e-35;
  Matches   152;  Conservative    0;  Mismatches  44;  Indels   3;  Gaps   3;

Db      3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
          ||| | ||||||||||| || ||||||||||||||  | ||||  |||||||||    |
Qy    993 AACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGG 1052

Db     63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
          ||  | |||||| ||    |||| || | ||   | | ||| |||   ||||||||||||
Qy   1053 TCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCAACCGCCGC 1109

Db    123 aagtgccacaaggcccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
          |||||||||||||||| || |||||||| |||  ||| ||   | || ||| |||| |
Qy   1110 AAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGA 1169

Db    183 atgctcttctgctcctgcc 201
          ||||||||||||||||||
Qy   1170 ATGCTCTTCTGCTCCTGCC 1188


RESULT    9
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            1.8%;  Score 45;  DB 1;  Length 204;
  Best Local Similarity  11.0%;  Pred. No. 1.07e-09;
  Matches    11;  Conservative   57;  Mismatches  31;  Indels   1;  Gaps   1;

Db     92 hhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnnc 151
          :: ::::: : ::: : :   ::  ::: ::::|| |  :::::::::::: :: :  |
Qy   1552 TTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCC-GATGTGACCGTGTGGCAGCCAGCC 1610

Db    152 ncccbnnhvchnvhbnnhrnwayvrhdarrddvhccvchc 191
            ||:  ::|: :::  :: : :::::|  : ::| :|:|
Qy   1611 TTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCC 1650


RESULT   10
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match            1.7%;  Score 44;  DB 9;  Length 91;
  Best Local Similarity  7.4%;  Pred. No. 3.52e-09;
  Matches     4;  Conservative   45;  Mismatches   5;  Indels   0;  Gaps   0;

Db     13 vhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctcaag 66
          :::::::::: :::::::::::::::::: :::::::::::: :::: | |||
Qy   1608 GCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAG 1661


RESULT   11
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
```

```
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             1.7%;  Score 42;  DB 9;  Length 91;
  Best Local Similarity   3.8%;  Pred. No. 3.72e-08;
  Matches     2;  Conservative   45;  Mismatches   5;  Indels   0;  Gaps   0;

Db     10 gssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvc 61
          |:::::::::::::::::: ::::::::: ::::::::::::::: :: : :|
Cp    617 GCGGTCTCCGCCGCTCACTTCGGCCGACAGGAGCAAGTCCAAGAGCGGCAGC 566


RESULT   12
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match             1.7%;  Score 43;  DB 1;  Length 204;
  Best Local Similarity  21.9%;  Pred. No. 1.15e-08;
  Matches    34;  Conservative   60;  Mismatches  60;  Indels   1;  Gaps   1;

Db     31 rccbgcaggycgacbcyrraggnyccccgggywccgagcycgaayycdchvgccgymrtt 90
          :||: || | ||  :| : |     ||  ||:  | || :|| |: | |::|| | ::||
Cp    772 GCCTCCATGGCGCTGCGGCACTCATCCTTGGCCTCCAGGCCGGATGC-CAGGCTGAAGTT 714

Db     91 thhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnn 150
            ::  : : :: :: ::: | : ||: ::::::|   ||::::: :::: :: :: :
Cp    713 GGTCTCCTTGCCCGCCACGCACTGCCTTAGCGTGCGGTACTTGGTGCTGCAGCTCTGCTC 654

Db    151 cncccbnnhvchnvhbnnhrnwayvrhdarrddvh 185
          | | :  ::    :    : : ::::: ::: ::
Cp    653 CTTCAGGCACTGATCACTGGCTTTCACGCAATCCA 619


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             1.5%;  Score 39;  DB 12;  Length 114;
  Best Local Similarity   7.2%;  Pred. No. 1.20e-06;
  Matches     8;  Conservative   32;  Mismatches  71;  Indels   0;  Gaps   0;

Db      1 tgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnn 60
          |||  :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :
Qy   1038 TGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCAACGATGTC 1097

Db     61 nnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
             :  : ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy   1098 TGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTC 1148


RESULT   14
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
```

```
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PT   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             1.5%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 3.76e-06;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnn 62
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|||  :
Qy     752 GTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCG 811

Db      63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnb 114
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :| |  :
Qy     812 GGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTACTGGAGCATGTACCAG 863


RESULT   15
ID   T76363 standard; DNA; 172 BP.
AC   T76363;
DT   15-SEP-1997  (first entry)
DE   Human interleukin 8 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 36; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human interleukin 8, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   172 BP;    0 A;    35 C;    42 G;    39 T;

  Query Match             1.5%;  Score 39;  DB 32;  Length 172;
  Best Local Similarity  33.8%;  Pred. No. 1.20e-06;
  Matches    45;  Conservative   41;  Mismatches  47;  Indels   0;  Gaps   0;

Db      14 bbbgcbtcbbgbbtbgctttgctbtctbbggbtcbcbtttbgbcbtbggbbbbcgctgtb 73
           :::  : |:: ::|:||   |      |:: |  |: :|| :|: :  ||:::|  || :
Cp     821 CTTCATACCCCGCTTGCAGCGGCAGTTGTAGAGCGACTTCTGCTTCAGGGCCTCCATGGC 762

Db      74 ggtcbgbbbgbtgtgcttbccttcbcbcbgbgctgcbgbbbtcbggbbggctgccbbgbg 133
           | | :|::  :  | |||: | ||: : : :| |||: :::|   |::||   ||::|:
Cp     761 GCTGCGGCACTCATCCTTGGCCTCCAGGCCGGATGCCAGGCTGAAGTTGGTCTCCTTGCC 702

Db     134 bgccbcggccbgc 146
           :||| ||  |:||
Cp     701 CGCCACGCACTGC 689


Search completed: Wed Jun 24 20:41:32 1998
Job time : 740 secs.
```

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jun 24 19:02:10 1998; MasPar time 2788.51 Seconds
1230.107 Million cell updates/sec
ular output not generated.

Title: >US-08-866-354-1
Description: (1-2568) from US08866354.seq
Perfect Score: 2525
N.A. Sequence: 1 AATCTGGCCTCGGAACACGC.....CACTCTTGTACAAAAGAAAA 2568
Comp: TTAGACCGGAGCCTTGTGCG.....GTGAGAACATGTTTTCTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1759237 seqs, 667866413 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est54
1:em_est1 2:em_est3
Database: genbank-est106
3:gb_est1 4:gb_est10 5:gb_est11 6:gb_est12 7:gb_est13
8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17
12:gb_est18 13:gb_est19 14:gb_est2 15:gb_est20
16:gb_est21 17:gb_est22 18:gb_est3 19:gb_est4 20:gb_est5
21:gb_est6 22:gb_est7 23:gb_est8 24:gb_est9 25:gb_gss
26:gb_sts

atistics: Mean 12.402; Variance 3.221; scale 3.850

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	495	19.6	515	7	AA512935	nh91b04.s1 NCI_CGAP_Br	0.00e+00
	2	428	17.0	631	7	AA518362	vi10d02.r1 Barstead mo	0.00e+00
	3	354	14.0	446	5	AA387098	vc18b07.r1 Ko mouse em	0.00e+00
	4	351	13.9	465	21	W53498	md45e01.r1 Soares mous	0.00e+00
	5	330	13.1	430	21	W76716	me81h06.r1 Soares mous	0.00e+00
c	6	303	12.0	330	9	AA573164	nm51f06.s1 NCI_CGAP_Br	0.00e+00
c	7	274	10.9	369	8	AA505897	ni01c09.s1 NCI_CGAP_Br	0.00e+00
	8	201	8.0	266	7	AA471183	PMY2139 KG1a Lambda Za	3.39e-254
	9	177	7.0	478	24	AA250120	mx21g08.r1 Soares mous	1.21e-216
c	10	158	6.3	487	13	AA780215	af52f08.s1 Soares tota	3.56e-187
	11	118	4.7	278	4	AA312082	EST182778 Jurkat T-cel	3.32e-126
	12	115	4.6	521	18	H12981	yl70a10.r1 Homo sapien	1.04e-121
	13	97	3.8	439	12	AA727460	vu96d10.r1 Stratagene	4.65e-95
	14	94	3.7	229	3	R02249	ye83h05.r1 Homo sapien	1.12e-90
	15	82	3.2	459	24	AA245689	mx03e07.r1 Soares mous	2.24e-73
c	16	66	2.6	252	13	AA754459	97SN1787 Rice Immature	5.19e-51
	17	62	2.5	252	13	AA754459	97SN1787 Rice Immature	1.38e-45
c	18	51	2.0	247	13	AA754458	97SN1784 Rice Immature	3.90e-31
	19	47	1.9	247	13	AA754458	97SN1784 Rice Immature	4.30e-26
c	20	43	1.7	180	23	AA064450	ml47h11.r1 Stratagene	3.39e-21
	21	42	1.7	350	4	AA349976	EST56965 Infant brain	5.35e-20
	22	38	1.5	2275	12	AF034173	Homo sapiens ntcon2 co	2.53e-15
	23	36	1.4	284	25	FR0012397	F.rubripes GSS sequenc	4.58e-13
	24	35	1.4	478	22	W73681	zd55h01.r1 Soares feta	5.85e-12
c	25	30	1.2	318	9	AA587886	nm71a08.s1 NCI_CGAP_Co	1.09e-06
	26	30	1.2	350	3	T09902	0455m3 Plasmodium falc	1.09e-06
c	27	30	1.2	360	25	B24410	F20A16TR IGF Arabidops	1.09e-06
c	28	30	1.2	434	12	AA083088	zn08h06.s1 Stratagene	1.09e-06
c	29	30	1.2	444	15	AA819896	UI-R-A0-aq-c-09-0-UI.s	1.09e-06
c	30	30	1.2	466	9	AA524108	ng35a11.s1 NCI_CGAP_Co	1.09e-06
c	31	30	1.2	524	10	AA427938	zw53d03.s1 Soares tota	1.09e-06
c	32	30	1.2	565	10	AA610066	af08e07.s1 Soares test	1.09e-06
c	33	31	1.2	2275	12	AF034173	Homo sapiens ntcon2 co	1.05e-07
c	34	29	1.1	298	26	HSB346YB5	H.sapiens (D1S2781) DN	1.07e-05
c	35	29	1.1	334	26	G10329	human STS CHLC.ATC6F11	1.07e-05
c	36	28	1.1	416	18	M78296	EST00444 Homo sapiens	1.00e-04
	37	29	1.1	432	3	T47812	yb17c05.r1 Homo sapien	1.07e-05
c	38	28	1.1	518	25	B67798	T24H12TR TAMU Arabidop	1.00e-04
c	39	28	1.1	522	7	C25513	Dictyostelium discoide	1.00e-04
c	40	29	1.1	655	25	B52556	CIT-HSP-2005L3.TR CIT-	1.07e-05
	41	29	1.1	777	25	AG004298	Homo sapiens genomic D	1.07e-05
	42	28	1.1	789	25	AG004299	Homo sapiens genomic D	1.00e-04
	43	29	1.1	808	25	AG004297	Homo sapiens genomic D	1.07e-05
c	44	29	1.1	814	25	AG004330	Homo sapiens genomic D	1.07e-05
	45	29	1.1	853	25	B19213	F10D23-T7 IGF Arabidop	1.07e-05

ALIGNMENTS

RESULT 1
LOCUS AA512935 515 bp mRNA EST 05-AUG-1997
DEFINITION nh91b04.s1 NCI_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:965839, mRNA sequence.
ACCESSION AA512935
NID g2251358
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 515)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1735 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 460.
FEATURES Location/Qualifiers
source 1. .515
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

```
                polylinker; 1st strand cDNA was prepared from pooled bulk
                breast tumor tissue, and was then primed with a Not I -
                oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                RI adaptors (Pharmacia), digested with Not I and cloned
                into the Not I and Eco RI sites of the modified pT7T3
                vector.  Library is not normalized. (The normalized
                version of this library is NCI_CGAP_Br2.) Library was
                constructed by Bento Soares and M. Fatima Bonaldo."
                /db_xref="taxon:9606"
                /clone="IMAGE:965839"
                /clone_lib="NCI_CGAP_Br1.1"
                /sex="female, pooled"
                /tissue_type="breast"
                /lab_host="DH10B"
     mRNA       <1. .>515
BASE COUNT      123 a    112 c    131 g    149 t
ORIGIN

  Query Match           19.6%;  Score 495;  DB 7;  Length 515;
  Best Local Similarity  98.8%;  Pred. No. 0.00e+00;
  Matches   511;  Conservative    0;  Mismatches   4;  Indels   2;  Gaps   2;

Db       1 TCTTTTTACATGTCCATATTGTATTTTTTTAATGCAGCTATGATGTTTCTGTTAAAGATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1974 TCTTTTTACATGTCCATATTGTATTTTTTTAATGCAGCTATGATGTTTCTGTTAAAGATA 1915

Db      61 ATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACAGCTTGGAG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1914 ATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACAGCTTGGAG 1855

Db     121 GAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAA-CACCGAGACCTTCTTTTTCATAAT 179
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Cp    1854 GAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAAGCACCGAGACCTTCTTTTTCATAAT 1795

Db     180 TACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAGCTTCTGTGCCT 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1794 TACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAGCTTCTGTGCCT 1735

Db     240 GTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGACCCTGCTGGCC 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1734 GTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGACCCTGCTGGCC 1675

Db     300 CCAGGGCTCTGTTCTTAACCCGGAGG-CAGTGGTGGTAGCGGCAGTGGTGGTCTGTACTG 358
           ||||||   ||||||||||||||||| |||||||||||| ||||||||||||||||||||
Cp    1674 CCAGGGGCTTGTTCTTAACCCGGAGGGCAGTGGTGGTAGTGGCAGTGGTGGTCTGTACTG 1615

      359 GGAAGGCTGGCTGCCACACGGTCACATCGGAGCCATTGCCAAAGGCTTGAATTGCATTTT 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1614 GGAAGGCTGGCTGCCACACGGTCACATCGGAGCCATTGCCAAAGGCTTGAATTGCATTTT 1555

Db     419 TAAGACATGTATTGTCCTTGAAGAAATTCAAAAATTTCAAGCACTCTTCTAGGTCGTTCC 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    1554 TAAGACATGTATTGTCCTTGAAGAAATTCAAAAATTTCAAGCACTCTTCTAGGTCGTTCC 1495

Db     479 CACTGTTGCTGCAGTCACACCATGGGGCCACACTGAG 515
           |||||||||||||||||||||||||||||||||||||
Cp    1494 CACTGTTGCTGCAGTCACACCATGGGGCCACACTGAG 1458


RESULT    2
LOCUS       AA518362      631 bp    mRNA            EST       16-JUL-1997
DEFINITION  vi10d02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            903363 5', mRNA sequence.
ACCESSION   AA518362
NID         g2259047
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:524027
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 259.
FEATURES              Location/Qualifiers
     source           1. .631
                      /organism="Mus musculus"
                      /strain="C3H"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I  and Eco RI sites of the modified pT7T3 vector.
                      Library constructed  by Bob Barstead.  The C2C12 cell line
                      (available from ATCC,  catalog # CRL-1772) differentiates
                      rapidly, forming contractile  myotubes and producing
                      characteristic muscle proteins."
                      /db_xref="taxon:10090"
                      /clone="903363"
                      /clone_lib="Barstead mouse myotubes MPLRB5"
                      /cell_line="C2C12"
                      /lab_host="DH10B"
     mRNA             <1. .>631
BASE COUNT      154 a    174 c    166 g    133 t      4 others
ORIGIN

  Query Match           17.0%;  Score 428;  DB 7;  Length 631;
  Best Local Similarity  87.6%;  Pred. No. 0.00e+00;
  Matches   516;  Conservative    0;  Mismatches  70;  Indels   3;  Gaps   3;

Db      14 GTGCCACAAAGCCCTCAGGCAGTTCTTCGACAAAGTACCAGCCAAGCACAGCTACGGGAT 73
           ||||||||| |||||| |||||||||| ||||| || || ||||||||||||||||| ||
Qy    1112 GTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAAT 1171

Db      74 GCTCTTCTGCTCCTGCCGGGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCC 133
           |||||||||||||||||||| |||||||||| ||| || |||||||||| ||||| ||
Qy    1172 GCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCC 1231

Db     134 TGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAA 193
           |||||||||||||||||  | |||| |||||||||  ||||| |||| ||||||||||
Qy    1232 TGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAA 1291

Db     194 GACAAATTACATCTGCAGATCTCGCCTTGCAGAGTTTTTTTACCAACTGCCAGCCAGAGT 253
           ||| |||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Qy    1292 GACGAATTACATCTGCAGATCTCGCCTTGCGGA-TTTTTTTACCAACTGCCAGCCAGAGT 1350

Db     254 CAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGG 313
           |||||||||||||||| |||||| ||||| |||||||| |||||||||| ||||||||||
Qy    1351 CAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGG 1410

Db     314 GACTGATTGGCACAGTCATGACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGC 373
           | || |||||||||||||||| || |||||||||||||||| |||||||||||||| |
Qy    1411 GGCTTATTGGCACAGTCATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCC 1470

Db     374 CGTGGTGCGATTGCAGCAGCAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATT 433
           | ||||| || ||||||| |||||| || ||||| ||||| ||| |||| ||| ||||||
```

```
Qy   1471 CATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATT 1530

Db    434 TTTTTAAGGGCAATACGTGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCGGATG 493
          | || |||| |||||| ||||| |||||||||||||||||||||||||||||||| ||||
Qy   1531 TCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATG 1590

Db    494 TGACCATGTGGCAGCCAGCAC-CACAGTCCAGACCACCAATGCCACGACTACCAATGCCT 552
          ||||| |||||||||||||   | |||| |||||||||| |||||| || |||| ||||
Qy   1591 TGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCC 1650

Db    553 TCGGGATCAAGAACGAGCTACTGGG-CCGGCAGGTTCTGAGAATGAGAT 600
          || || | |||||| ||| |||||| || ||||| |||||||||| ||
Qy   1651 TCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTGAGAATGAAAT 1699


RESULT    3
LOCUS       AA387098      446 bp    mRNA            EST       23-APR-1997
DEFINITION  vc18b07.r1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone 774901
            5', mRNA sequence.
ACCESSION   AA387098
NID         g2040052
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:467757
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total
                     RNAs were extracted from 11.5 dpc embryos (excluding
                     placenta and yolk sac). The double-stranded cDNA was
                     synthesized with an oligo (dT)-1 primer
                     GAGAGAGACTAGTTCTAGATCGCGAGCGGCCGCTTTTTTTTTTTTTTTT 3'.
                     The cDNAs were ligated to LL-Sal3A: 5'
                     GCTATTGACGTCGACTATCC 3' and LL-Sal3B: 5'
                     GGATAGTCGACGTCAAT 3'. The cDNAs were size-selected and
                     amplified by long-range PCR using Ex Taq polymerase for 18
                     cycles. The PCR-amplifiable cDNA mixture went through
                     one round of equalization and was digested with SalI/NotI
                     and cloned into the SalI/NotI sites of the pSPORT1
                     plasmid vector (Life Technologies). The library was
                     constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong
                     Wang."
                     /db_xref="taxon:10090"
                     /clone="774901"
                     /clone_lib="Ko mouse embryo 11 5dpc"
                     /sex="pooled"
                     /dev_stage="11.5dpc"
                     /lab_host="DH10B"
     mRNA            <1. .>446
```

```
BASE COUNT      112 a    122 c    107 g    105 t
ORIGIN

  Query Match          14.0%;  Score 354;  DB 5;  Length 446;
  Best Local Similarity  89.7%;  Pred. No. 0.00e+00;
  Matches   400;  Conservative    0;  Mismatches  46;  Indels   0;  Gaps   0;

Db      1 GCAGTTCTTCGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 60
          |||||||| ||||| || || |||||||||||||||| ||||||||||||||||||||
Qy   1130 GCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCG 1189

Db     61 GGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAGA 120
          |||| |||||||||| ||| || |||||||||| ||||| ||||||||||||||||||||
Qy   1190 GGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGA 1249

Db    121 ACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAG 180
            | |||| |||||||||  ||||| |||| ||||||||||||||| ||||||||||||||
Qy   1250 GAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAG 1309

Db    181 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 240
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||| |||
Qy   1310 ATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTG 1369

Db    241 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 300
          ||| ||||| |||||||| |||||||||| |||||||||| || ||||||||||||||
Qy   1370 TCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCAT 1429

Db    301 GACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAA 360
          ||| || |||||||||||||||| |||||||||||||| || ||||| || ||||||||
Qy   1430 GACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAA 1489

Db    361 CAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTG 420
          |||||| || ||||| ||||| ||| |||| ||| ||||||| || |||||||||| ||
Qy   1490 CAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATG 1549

Db    421 TCTCAAAAATGCAATTCAAGCCTTTG 446
          ||| |||||||||||||||||||||||
Qy   1550 TCTTAAAAATGCAATTCAAGCCTTTG 1575


RESULT    4
LOCUS       W53498        465 bp    mRNA            EST       03-JUN-1996
DEFINITION  md45e01.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 371352 5', mRNA sequence.
ACCESSION   W53498
NID         g1357323
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:232784
            Seq primer: mob.REGA+ET
            High quality sequence stop: 347.
```

```
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="371352"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
     mRNA            <1. .>465
BASE COUNT      120 a    133 c    103 g    109 t
ORIGIN

  Query Match           13.9%;  Score 351;  DB 21;  Length 465;
  Best Local Similarity  89.0%;  Pred. No. 0.00e+00;
  Matches   412;  Conservative    0;  Mismatches  49;  Indels   2;  Gaps   2;

Db       4 TGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATTTTT 63
           ||||| |||| |||||||||||||| ||||||||||||||||||||||||||| |||||||
Qy    1270 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 1329

Db      64 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAG 123
           |||||||||||||||||||||||||||||||||| |||||| ||||| |||||||| |
Qy    1330 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 1389

Db     124 ACTGCCTCCT-GCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAGACT 182
           |||||||||| |||||||||| || |||||||||||||||| || ||||||||||||
Qy    1390 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1449

Db     183 CCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGGAAG 242
           |||| |||||||||||||| || ||||| || |||||||||||||| || ||||| ||||
Qy    1450 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1509

       243 ATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTCAAG 302
           | ||| |||| ||| ||||||| || |||||||||| ||||| ||||||||||||||||
Qy    1510 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1569

Db     303 CCTTTGCACATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCAGTCCAGACCACCA 362
           ||||||   ||||||| ||||||||| |||||||||||||||  |||||| ||||||||||
Qy    1570 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1629

Db     363 CTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCTCTG 422
           ||||||| || |||||||| ||||| | |||||||||| || |||||||||| ||||
Qy    1630 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1689

Db     423 AGAATGAGATTCC-ACACACGTTTTACCACCGTGTGCTAATTT 464
           ||||||| ||||| || || ||||| |||||||||| |||||
Qy    1690 AGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTT 1732


RESULT    5
LOCUS       W76716         430 bp    mRNA            EST       20-JUN-1996
DEFINITION  me81h06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 402011 5', mRNA sequence.
ACCESSION   W76716
NID         g1387762
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:245779
            Seq primer: mob.REGA+ET
            High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="402011"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
     mRNA            <1. .>430
BASE COUNT      112 a    121 c    100 g     97 t
ORIGIN

  Query Match           13.1%;  Score 330;  DB 21;  Length 430;
  Best Local Similarity  89.2%;  Pred. No. 0.00e+00;
  Matches   381;  Conservative    0;  Mismatches  45;  Indels   1;  Gaps   1;

Db       4 TGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATTTTT 63
           ||||| |||| |||||||||||||| ||||||||||||||||||||||||||| |||||||
Qy    1270 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 1329

Db      64 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAG 123
           |||||||||||||||||||||||||||||||||| |||||| ||||| |||||||| |
Qy    1330 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 1389

Db     124 ACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAGACT 183
           |||||||||| |||||||||| || |||||||||||||||| || ||||||||||||||
Qy    1390 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1449

Db     184 CCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGGAAG 243
           |||| |||||||||||||| || ||||| || |||||||||||||| || ||||| ||||
Qy    1450 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1509

Db     244 ATTGCCTGAAGTTTCTGAATTTTTTTAAGGA-AATACGTGTCTCAAAAATGCAATTCAAG 302
           | ||| |||| ||| ||||||| || ||||| ||||| ||||| ||||||||||||||||
```

```
Qy   1510 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1569

Db    303 CCTTTGCAGATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCCAGTCCAGACCACCA 362
          ||||||   ||||||| |||||||||  ||||||||||||||  ||||||  ||||||||||
Qy   1570 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1629

Db    363 CTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCTCTG 422
          ||||||| || |||||||||  |||||  | ||||||||||| ||  |||||||||| ||||
Qy   1630 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1689

Db    423 AGAATGA 429
          |||||||
Qy   1690 AGAATGA 1696
```

```
RESULT    6
LOCUS       AA573164      330 bp    mRNA            EST       28-AUG-1997
DEFINITION  nm51f06.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1071779, mRNA
            sequence.
ACCESSION   AA573164
NID         g2347692
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1071779"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
     mRNA            <1. .>330
BASE COUNT       75 a     62 c     75 g    118 t
ORIGIN

  Query Match           12.0%;  Score 303;  DB 9;  Length 330;
  Best Local Similarity  99.4%;  Pred. No. 0.00e+00;
  Matches   310;  Conservative    0;  Mismatches   1;  Indels   1;  Gaps   1;

Db     20 TTTTTTTCTTTTTACATGTCCATATTGTATTTTTTTAATGCAGCTATGATGTTTCTGTTA 79
```

```
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1980 TTTTTGTCTTTTTACATGTCCATATTGTATTTTTTTAATGCAGCTATGATGTTTCTGTTA 1921

Db     80 AAGATAATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACAGC 139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1920 AAGATAATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACAGC 1861

Db    140 TTGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAA-CACCGAGACCTTCTTTTT 198
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Cp   1860 TTGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAAGCACCGAGACCTTCTTTTT 1801

Db    199 CATAATTACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAGCTTCT 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1800 CATAATTACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAGCTTCT 1741

Db    259 GTGCCTGTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGACCCTG 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp   1740 GTGCCTGTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGACCCTG 1681

Db    319 CTGGCCCCAGGG 330
          ||||||||||||
Cp   1680 CTGGCCCCAGGG 1669
```

```
RESULT    7
LOCUS       AA505897      369 bp    mRNA            EST       18-AUG-1997
DEFINITION  ni01c09.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:966736, mRNA
            sequence.
ACCESSION   AA505897
NID         g2242034
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:966736"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
```

```
                  /lab_host="DH10B"
     mRNA         <1. .>369
BASE COUNT      126 a     69 c     67 g    107 t
ORIGIN

  Query Match           10.9%;  Score 274;  DB 8;  Length 369;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   274;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      16 TTTTCTTTTGTACAAGAGTGCAGAACTGTTTACAGTACATAAACATCAGCGTGGAAAATG 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2568 TTTTCTTTTGTACAAGAGTGCAGAACTGTTTACAGTACATAAACATCAGCGTGGAAAATG 2509

Db      76 CTAGTACATATATAAAGGTGAACATCAGTACCAGCTGTTTCCACAAAGCCCACTCTTGGT 135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2508 CTAGTACATATATAAAGGTGAACATCAGTACCAGCTGTTTCCACAAAGCCCACTCTTGGT 2449

Db     136 AAGAATCTTCTTTGTGGCAAAAAAGCTTGGCATCAATGTAGGCTGAAATTTATATCATTA 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2448 AAGAATCTTCTTTGTGGCAAAAAAGCTTGGCATCAATGTAGGCTGAAATTTATATCATTA 2389

Db     196 GAAATCAATGTCATTAAAGGAGATATGGGAGTTACCTTAGAAATATTAACTCCTTCTCTA 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Cp    2388 GAAATCAATGTCATTAAAGGAGATATGGGAGTTACCTTAGAAATATTAACTCCTTCTCTA 2329

Db     256 GCTAGGAAAGGCCAGAAGTAAAACTGTTAAAATC 289
           |||||||||||||||||||||||||||||||||
Cp    2328 GCTAGGAAAGGCCAGAAGTAAAACTGTTAAAATC 2295


RESULT    8
LOCUS       AA471183      266 bp    mRNA            EST       17-JUN-1997
DEFINITION  PMY2139 KG1a Lambda Zap Express cDNA Library Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA471183
NID         g2199040
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Claudio,J.O., Liew,C.C. and Hawley,R.G.
  TITLE     Hematopoietic progenitor ESTs
  JOURNAL   Unpublished (1997)
  MENT
            Contact: Hawley RG
            Oncology Research Laboratories
            The Toronto Hospital
            CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
            Tel: 416 3403834
            Fax: 416 3403453
            Email: r.hawley@utoronto.ca
            Clone was randomly picked from subtracted KG1a primary cDNA
            library.
            Seq primer: 5' GAAATTAACCCTCACTAAAGGG 3'
            High quality sequence stop: 226.
FEATURES              Location/Qualifiers
     source           1. .266
                      /organism="Homo sapiens"
                      /note="Vector: Lambda Zap Express (Stratagene); Site_1:
                      EcoRI; Site_2: XhoI; Unidirectional cloning sites:
                      EcoRI-XhoI. mRNA was purified from KG1a cell line, cDNA
                      was synthesized using an XhoI-OligodT linker primer.
                      EcoRI adaptors were ligated, followed by digestion with
                      XhoI for directional cloning into predigested Lambda Zap
                      Express."
                      /db_xref="taxon:9606"
                      /clone_lib="KG1a Lambda Zap Express cDNA Library"
                      /cell_type="promyeloblast"
                      /cell_line="KG1a"
     mRNA             <1. .>266
```

```
BASE COUNT       70 a     74 c     57 g     65 t
ORIGIN

  Query Match           8.0%;  Score 201;  DB 7;  Length 266;
  Best Local Similarity 95.5%;  Pred. No. 3.39e-254;
  Matches   253;  Conservative    0;  Mismatches   4;  Indels   8;  Gaps   8;

Db       1 TCTTAAAAATGCAATTCAAGCCTTTTGCAATGTGCTCCGATGTGACCGTGTGGCAGCCAG 60
           |||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||
Qy    1550 TCTTAAAAATGCAATTCAAGCCTTTGGCAATG-GCTCCGATGTGACCGTGTGGCAGCCAG 1608

Db      61 CCTTCCCAGTACAGACCACCACTGCCCACTACCCACCACTTGCCCTTCCGGGTTAAGAAC 120
           |||||||||||||||||||||||| |||||| |||||| ||||| ||||||||||||||
Qy    1609 CCTTCCCAGTACAGACCACCACTGCC-ACTACC-ACCACT-GCCCT-CCGGGTTAAGAAC 1664

Db     121 AAGCCCCTTGGGGCCAGCAGGGGTCTGAGAATGAAATTTCCACTCATGTTTTGCCACCGT 180
           |||||||| |||||||||||||| |||||||||||||| ||||||||||||||||||||
Qy    1665 AAGCCCCT-GGGGCCAGCAGGG-TCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGT 1722

Db     181 GTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCA-TACACACCTCTGTA 239
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Qy    1723 GTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTA 1782

Db     240 TTTCCACTGGTAATTATGTAAAAGA 264
           |||||| |||||||||||| ||||||
Qy    1783 TTTCCAATGGTAATTATGAAAAAGA 1807


RESULT    9
LOCUS       AA250120      478 bp    mRNA            EST       12-MAR-1997
DEFINITION  mx21g08.r1 Soares mouse NML Mus musculus cDNA clone 680894 5', mRNA
            sequence.
ACCESSION   AA250120
NID         g1882418
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:420598
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
```

```
                     /db_xref="taxon:10090"
                     /clone="680894"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
     mRNA            <1. .>478
BASE COUNT      121 a    116 c    111 g    130 t
ORIGIN

  Query Match           7.0%;  Score 177;  DB 24;  Length 478;
  Best Local Similarity  75.7%;  Pred. No. 1.21e-216;
  Matches   370;  Conservative    0;  Mismatches 104;  Indels  15;  Gaps   9;

Db       1 GCTAATTTGCAGGCACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTT 60
          || ||||| |||||||||||||||||||||||||| ||||||| |||||| |||||| ||
Qy    1725 GCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATT 1784

Db      61 TCTGATAATGATTACGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAA 120
          ||  ||  | |||| | ||| || |||||||   |||| ||||||||||||||||||||
Qy    1785 TCCAATGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAA 1841

Db     121 TCAATGGCTGCTCCTCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCT 180
          |||||||||||||||| ||||| ||||||||| |||||| |||  ||||| | ||||||
Qy    1842 TCAATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCT 1901

Db     181 CTGGCTGCCCTGTTGTCTGTATCATTGGCAGAA-ACAT-CGT-AGCTGCATCCAGTATGA 237
          ||| |  |||| || ||| || ||    ||  | |  | | | |     || || ||||
Qy    1902 CTGTCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGG 1961

Db     238 AAA-GACAAAAGA---AAACCAAGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATC 293
          | | |  ||||||   |||||||||  |||||  ||||| |||||||||| ||||||
Qy    1962 ACATGTAAAAAGACAAAAACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTC 2021

Db     294 CAGTTTTAAAAGCTCGGTTGAGAAGCAGTTTCACCCAGCGGGAAC-TCCTTTGTTGT--- 349
          |||||| |  ||||| |||||||| ||||| ||  || | ||||| |  ||| || |
Qy    2022 CAGTTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCC 2080

Db     350 TTTTAAGAAAGCTTTTTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAA 409
          |||||||||||||| |||    |||    |||||||||  || | ||| | ||| |||
Qy    2081 TTTTAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCC 2140

Db     410 TTCCAAACTCAGAAAACTCTGGGGCATGGTGTGGCTTCAGGGGACTGTTTGTAACCTGGA 469
           |||||||||||||  || ||||  ||| |||   || | ||||| |||||||| |||
Qy    2141 ATCCAAACTCAGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGG 2200

Db     470 CTGTAAAGC 478
           |||||||||
Qy    2201 CTGTAAAGC 2209
```

```
RESULT   10
LOCUS       AA780215      487 bp    mRNA            EST       05-FEB-1998
DEFINITION  af52f08.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
            1035303 3', mRNA sequence.
ACCESSION   AA780215
NID         g2839546
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 461.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3']."
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="1035303"
                     /clone_lib="Soares total fetus Nb2HF8 9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
BASE COUNT      169 a     78 c     93 g    147 t
ORIGIN

  Query Match           6.3%;  Score 158;  DB 13;  Length 487;
  Best Local Similarity  97.7%;  Pred. No. 3.56e-187;
  Matches   167;  Conservative    0;  Mismatches   3;  Indels   1;  Gaps   1;

Db     317 TTTACTTTTGTAAAAGAGTGCAGAACTGTTTACAGTACATAAACATCAGCGTGGAAAATG 376
           ||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Cp    2568 TTTTCTTTTGTACAAGAGTGCAGAACTGTTTACAGTACATAAACATCAGCGTGGAAAATG 2509

Db     377 CTAGTACATATATAAAGGTGAACATCAGTACCAGCTGTTTCCACAAAGCCCACTACTTGG 436
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Cp    2508 CTAGTACATATATAAAGGTGAACATCAGTACCAGCTGTTTCCACAAAGCCCACT-CTTGG 2450

Db     437 TAAGAATCTTCATTGTGGCAAAAAAGCTTGGCATCAATGTAGGCTGAAATT 487
           |||||||||| ||||||||||||||||||||||||||||||||||||||||
Cp    2449 TAAGAATCTTCTTTGTGGCAAAAAAGCTTGGCATCAATGTAGGCTGAAATT 2399
```

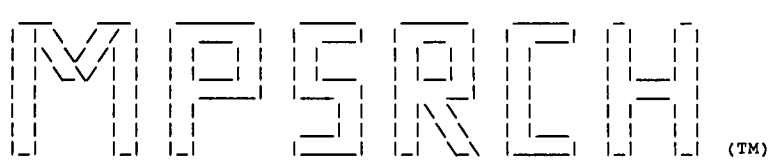
```
RESULT   11
LOCUS       AA312082      278 bp    mRNA            EST       19-APR-1997
DEFINITION  EST182778 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA312082
NID         g1964409
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 278)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
```

```
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .278
                      /organism="Homo sapiens"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI"
                      /db_xref="ATCC (inhost):159272"
                      /db_xref="taxon:9606"
                      /clone_lib="Jurkat T-cells VI"
                      /cell_type="T-lymphocyte"
     mRNA             <1. .>278
BASE COUNT       51 a     89 c     90 g     42 t      6 others
ORIGIN

  Query Match             4.7%;  Score 118;  DB 4;  Length 278;
  Best Local Similarity  99.2%;  Pred. No. 3.32e-126;
  Matches   118;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    160 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGNTTCCAATAACCACTAACATCCCTAA 219
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Qy      1 AATCTGGCCTCGGAACACGCCATTCTCCGCGCCGCTTCCAATAACCACTAACATCCCTAA 60

Db    220 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGA 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 CGAGCATCCGAGCCGAGGGCTCTGCTCGGAAATCGTCCTGGCCCAACTCGGCCCTTCGA 119


RESULT   12
LOCUS       H12981        521 bp    mRNA            EST       27-JUN-1995
DEFINITION  yl70a10.r1 Homo sapiens cDNA clone 43207 5'.
ACCESSION   H12981
NID         g877801
KEYWORDS    EST.
SOURCE      human clone=43207 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-415-748
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 278
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .521
                      /organism="Homo sapiens"
                      /clone="43207"
BASE COUNT       95 a    154 c    144 g    119 t      9 others
ORIGIN

  Query Match             4.6%;  Score 115;  DB 18;  Length 521;
  Best Local Similarity  74.9%;  Pred. No. 1.04e-121;
  Matches   209;  Conservative    0;  Mismatches  65;  Indels   5;  Gaps   5;

Db      2 AACCATTGCCTGGGATGCTGCCAAGGCCTGCAACCTGAATGACAACTGCAAGAAGCTGCG 61
          ||| | ||||||| |||| || ||||||||||||||  | ||||  |||||||||    |
Qy    993 AACAACTGCCTGG-ATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAG 1051

Db     62 CTCCTCCTACATCTCCATCTGCAACCGCGAGATCTCGCCCACCGAGCGCTGCAACCGCCG 121
            ||  | |||||| ||   |||| || | ||   | | ||| |||   ||||||||||||
Qy   1052 GTCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCAACCGCCG 1108

Db    122 CAAGTGCCACAAGGCCCTGCGCCAGTTCTTCGACCGGGTGCCCAGCGAGTACACCTACCG 181
          |||||||||||||||||| || |||||||| |||  ||| ||   | || ||| |||| |
Qy   1109 CAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGG 1168

Db    182 CATGCTCTTCTGCTCCTGCCAAGACCAGGCGTGCGCTGAGCCGCGGNCAAA-ACCATCCT 240
          ||||||||||||||||||  |||   || ||| | |||| | ||  | | |||||| |
Qy   1169 AATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGT 1228

Db    241 GCCCAGCTGCTCCTATGAGGACAAGGAGAAGCCCAACTG 279
          |||    |||||||||||| || | ||||||||||||||||
Qy   1229 GCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTG 1267


RESULT   13
LOCUS       AA727460      439 bp    mRNA            EST       02-JAN-1998
DEFINITION  vu96d10.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1210003 5', mRNA sequence.
ACCESSION   AA727460
NID         g2745167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```

```
            MGI:646347
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 436.
FEATURES              Location/Qualifiers
     source           1. .439
                      /organism="Mus musculus"
                      /strain="C57BL/6"
                      /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="1210003"
                      /clone_lib="Stratagene mouse skin (#937313)"
                      /sex="females"
                      /tissue_type="whole skin"
                      /dev_stage="11 weeks old"
                      /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      105 a     79 c    117 g    138 t
ORIGIN

  Query Match             3.8%;  Score 97;  DB 12;  Length 439;
  Best Local Similarity  73.6%;  Pred. No. 4.65e-95;
  Matches   201;  Conservative    0;  Mismatches  65;  Indels   7;  Gaps   4;

Db      1 AGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAGGTTTTAAAAGCTCGGTTGA 60
          |||  |||||  |||||  |||||||||||| ||||||  |||||  ||||  |||||  |||||
Qy   1984 AGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAG--TTTAGGAGCTCAGTTGA 2041

Db     61 GAAGCAGTTTCACCCAGC-GGAAC-TCCTTTGTTGT---TTTTAAGAAAGCTTTTTGCCG 115
          ||| |||||  ||  || | |||||  |  |||  || |   ||||||||||||||  |||
Qy   2042 GAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTNCCTTTTAAGAAAGCTTCTTGTGA 2101

Db    116 CCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACTCAGAAAACTCTG 175
           |||    |||||||||  || | ||| | ||| |||   ||||||||||||  || ||
Qy   2102 TCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAACTCAGAAGGCTTTG 2161

Db    176 GGGCATGGTGTGGCTTCAGGGGACTGTTTGTAACCTGGACTGTAAAGCAAGCTGAGCTTA 235
          ||  ||| |||   || | ||||| |||||||| ||| ||||||||||| ||| |  |
Qy   2162 GGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAAGCAAACTGGGGCTG 2221

Db    236 TGTTTTTGATGGTGATGGTGGTGGTGATGATGA 268
          |||||| |||| ||||| |  |  |||| ||||
Qy   2222 TGTTTTCGATGATGATGATNATCATGATNATGA 2254


RESULT   14
LOCUS       R02249        229 bp    mRNA            EST       31-MAR-1995
DEFINITION  ye83h05.r1 Homo sapiens cDNA clone 124377 5'.
ACCESSION   R02249
NID         g751985
KEYWORDS    EST.
SOURCE      human clone=124377 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsite1=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 214
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .229
                      /organism="Homo sapiens"
                      /clone="124377"
BASE COUNT       49 a     85 c     55 g     39 t      1 others
ORIGIN

  Query Match             3.7%;  Score 94;  DB 3;  Length 229;
  Best Local Similarity  74.0%;  Pred. No. 1.12e-90;
  Matches   168;  Conservative    0;  Mismatches  56;  Indels   3;  Gaps   3;

Db      1 CCAAGAGCAACCATTGCCTGGATGCTGCCAAGGCCTGCAACCTGAATGACAACTGCAAGA 60
          ||||  | ||| | |||||||||||| || ||||||||||||||  | ||||  ||||||
Qy    985 CCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGA 1044

Db     61 AGCTGCGCTCCTCCTACATCTCCATCTGCAACCGCGAGATCTCGCCCACCGAGCGCTGCA 120
          ||    | ||  | |.|||| ||   |||| || | ||   | | ||| |||   |||||
Qy   1045 AGTACAGGTCGGCGTACATCACCCCGTGCA-CCACCAGCG-T-GTCCAACGATGTCTGCA 1101

Db    121 ACCGCCGCAAGTGCCACAAGGCCCTGCGCCAGTTCTTCGACCGGGTGCCCAGCGAGTACA 180
          |||||||||||||||||||||||||| || |||||||| |||  ||| ||   | || |||
Qy   1102 ACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACA 1161

Db    181 CCTACCGCATGCTCTTCTGCTCCTGCCAAGATCAGGCGTGCGCTGAG 227
          |||| | ||||||||||||||||||  ||    || ||| | |||
Qy   1162 GCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAG 1208


RESULT   15
LOCUS       AA245689      459 bp    mRNA            EST       10-MAR-1997
DEFINITION  mx03e07.r1 Soares mouse NML Mus musculus cDNA clone 679140 5', mRNA
            sequence.
ACCESSION   AA245689
NID         g1876478
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:418844
```

```
           Seq primer: -28m13 rev2 ET from Amersham
           High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="679140"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
     mRNA            <1. .>459
BASE COUNT      117 a     87 c    118 g    137 t
ORIGIN

  Query Match             3.2%;  Score 82;  DB 24;  Length 459;
  Best Local Similarity  73.1%;  Pred. No. 2.24e-73;
  Matches   125;  Conservative    0;  Mismatches  46;  Indels   0;  Gaps   0;

Db      1 TAAGAAAGCTTTTTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTC 60
          |||||||||||  |||    |||    |||||||||  || | ||| | ||| |||   ||
Qy   2084 TAAGAAAGCTTCTTGTGATCCTTNGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATC 2143

Db     61 CAAACTCAGAAAACTCTGGGGCATGGTGTGGCTTCAGGGGACTGTTTGTAACCTGGACTG 120
          ||||||||||  || ||||  ||| |||   || | ||||| |||||||| ||| |||
Qy   2144 CAAACTCAGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTGGGCTG 2203

Db    121 TAAAGCAAGCTGAGCTTATGTTTTTGATGGTGATGGTGGTGGTGATGATGA 171
          ||||||||  ||| |  | ||||||  |||| |||||  |  |  ||||  ||||
Qy   2204 TAAAGCAAACTGGGGCTGTGTTTTCGATGATGATGATNATCATGATNATGA 2254


Search completed: Wed Jun 24 20:28:51 1998
Job time : 5201 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 23 18:25:59 1998; MasPar time 15.62 Seconds
456.576 Million cell updates/sec
ular output not generated.

Title: >US-08-866-354-2
Description: (1-465) from US08866354.pep
Perfect Score: 3369
Sequence: 1 MFLATLYFALPLLDLLLSAE.....PLLVLVVTALSTLLSLTETS 465

Scoring table: PAM 150
Gap 11

Searched: 124785 seqs, 15338987 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27

Statistics: Mean 34.730; Variance 143.021; scale 0.243

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3184	94.5	468	27	W27327	Rat glial cell derive	0.00e+00
2	96	2.8	505	1	P80885	Sequence encoded by L	1.61e+01
3	87	2.6	32	3	R15252	Carbohydrate binding	7.18e+01
4	86	2.6	119	27	W43305	p17 antigen of human	8.45e+01
5	86	2.6	120	2	P70272	The sequence encoding	8.45e+01
6	86	2.6	132	5	R26326	HTLV-IIIb p17 sequenc	8.45e+01
7	86	2.6	169	10	R55851	HIV-1 MA delta-5-16 m	8.45e+01
8	86	2.6	171	10	R55857	HIV-1 MA delta-105-11	8.45e+01
9	86	2.6	267	2	P94509	Sequence encoded by p	8.45e+01
10	86	2.6	446	22	W19681	ATM mutant C1339T.	8.45e+01
11	86	2.6	478	3	P60309	Sequence of gag and p	8.45e+01
12	87	2.6	504	3	P93707	Sequence of the gag p	7.18e+01
13	86	2.6	506	6	R29704	gag gene decoded from	8.45e+01
14	86	2.6	512	2	R10866	HIV gag precursor pro	8.45e+01
15	87	2.6	514	3	R15237	Fusarium oxysporum C-	7.18e+01
16	86	2.6	521	23	W13053	HIV-2 provirus-encode	8.45e+01
17	88	2.6	522	2	R04024	Gag gene product of c	6.10e+01
18	86	2.6	524	27	W43304	Modified p17 antigen	8.45e+01
19	86	2.6	563	22	W19680	ATM mutant 1563delAG.	8.45e+01
20	86	2.6	600	2	P70541	HTLV-III gag/env gene	8.45e+01
21	88	2.6	607	7	R34662	Rat 5HT transporter.	6.10e+01
22	86	2.6	721	22	W19678	ATM mutant 2125del125	8.45e+01
23	86	2.6	721	22	W19687	ATM mutant 2125del125	8.45e+01
24	86	2.6	733	22	W19679	ATM mutant 2113delT.	8.45e+01
25	88	2.6	947	27	W42402	Amino acid sequence o	6.10e+01
26	86	2.6	954	22	W19674	ATM mutant 2839del83.	8.45e+01
27	87	2.6	1146	3	R15156	Abelson Related Gene,	7.18e+01
28	86	2.6	1407	22	W19672	ATM mutant 4110del127	8.45e+01
29	86	2.6	1483	22	W19671	ATM mutant 4437del175	8.45e+01
30	86	2.6	1625	22	W19689	ATM mutant 4777del830	8.45e+01
31	86	2.6	1789	22	W19668	ATM mutant 5320del7.	8.45e+01
32	86	2.6	1855	22	W19665	ATM mutant 5554insC.	8.45e+01
33	86	2.6	1928	22	W19664	ATM mutant 5712insA.	8.45e+01
34	86	2.6	2193	22	W19660	ATM mutant 6573del5.	8.45e+01
35	86	2.6	2652	22	W19655	ATM mutant 7883del5.	8.45e+01
36	86	2.6	2759	22	W19652	ATM mutant 8269del403	8.45e+01
37	86	2.6	2766	22	W19651	ATM mutant 8283delTC.	8.45e+01
38	86	2.6	2998	22	W19673	ATM mutant 3403del174	8.45e+01
39	86	2.6	2998	22	W19691	ATM mutant 3403del174	8.45e+01
40	86	2.6	3001	22	W19670	ATM mutant 4612del165	8.45e+01
41	86	2.6	3026	22	W19676	ATM mutant 2377del90.	8.45e+01
42	86	2.6	3054	22	W19695	ATM mutant 7279del6.	8.45e+01
43	86	2.6	3056	23	W19733	Cell cycle checkpoint	8.45e+01
44	86	2.6	3056	27	W06234	ATM protein.	8.45e+01
45	86	2.6	3056	26	W07655	Ataxia-telangiectasia	8.45e+01

ALIGNMENTS

```
RESULT    1
ID   W27327 standard; Protein; 468 AA.
AC   W27327;
DT   27-APR-1998  (first entry)
DE   Rat glial cell derived neurotrophic factor receptor alpha.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; rat; kidney disease; glomerulonephritis; therapy.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   Peptide          1..24
FT                    /label= Sig_peptide
FT   Domain           25..445
FT                    /note= "extracellular domain"
FT   Domain           446..468
FT                    /note= "GPI-attachment site"
FT   Modified_site    61
FT                    /note= "O-glycosylated"
FT   Modified_site    349
FT                    /note= "O-glycosylated"
FT   Modified_site    408
FT                    /note= "O-glycosylated"
FT   Cleavage_site    428..430
FT                    /note= "GPI-anchor cleavage/attachment site"
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
DR   N-PSDB; T84975.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Claim 1; Page 78-79; 100pp; English.
CC   This protein comprises full-length rat glial cell derived
CC   neurotrophic factor receptor alpha (GDNFR alpha), a novel
CC   GPI-linked protein that is a ligand-binding component of the
CC   receptor system for GDNF.  Its amino acid sequence was deduced
CC   from an isolated cDNA clone (see T84975).  The invention relates
CC   to novel uses of GDNF and its receptor.  In particular, it
CC   relates to native rat GDNFR alpha (see W27327), its variants and
```

```
CC   soluble derivatives (extracellular domain), chimeric GDNFR alpha
CC   and antibodies which bind to the GDNFR alpha, including agonist
CC   and neutralising antibodies, as well as various uses for these
CC   molecules.  It also relates to assay systems for detecting ligands
CC   to GDNFR alpha, systems for studying the physiological role of
CC   GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC   methods for identifying molecules homologous to GDNFR alpha, and
CC   therapeutic techniques (claimed) for the treatment of GDNF-related
CC   and GDNFR alpha-related conditions, particularly kidney disease
CC   associated with glomerulonephritis and enteric nervous system
CC   related disorders.  Transgenic and knockout animals are also
CC   claimed.
SQ   Sequence   468 AA;

  Query Match           94.5%;  Score 3184;  DB 27;  Length 468;
  Best Local Similarity  92.9%;  Pred. No. 0.00e+00;
  Matches   430;  Conservative   27;  Mismatches   5;  Indels   1;  Gaps   1;

        1 mflatlyfalplldllmsaevsggdrldcvkasdqclkeqscstkyrtlrqcvagketnf 60
          ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
        1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 sltsgleakdecrsamealkqkslyncrckrgmkkeknclriywsmyqslqgndlledsp 120
           ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 yepvnsrlsdifravpfisdvfqqvehiskgnncldaakacnlddtckkyrsayitpctt 180
           |||||||||||||| ||||||||||||||:|||||||||||||||| ||||||||||||||
Qy     121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db     181 smsnevcnrrkchkalrqffdkvpakhsygmlfcscrdiacterrrqtivpvcsyeerer 240
           |:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Qy     181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db     241 pnclslqdscktnyicrsrladfftncqpesrsvsnclkenyadcllaysgligtvmtpn 300
           ||||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Qy     241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db     301 yvdssslsvapwcdcsnsgndledclkflnffkdntclknaiqafgngsdvtmwqpappv 360
           |:|||||||||||||||||||:|||||||||||||||||||||||||||:|||| ||
Qy     301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db     361 qtttatttfafrvknkplgpagseneipthvlppcanlqaqklksnvsgsthlclsdsdf 420
           ||||||||||:|||||||||||||||||||||||||||||||||||||:||||:|::::
Qy     361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

       421 gkdglagasshittksmaappscslsslpvlmltalaallsvs 463
             |:|| ||||||||||||||||:||:| ||::|||::|||::
       421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462


RESULT    2
ID   P80885 standard; protein; 505 AA.
AC   P80885;
DT   16-DEC-1990  (first entry)
DE   Sequence encoded by LAV MA L GAG gene
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
OS   Lymphadenopathy associated virus MA L.
PN   WO8707906-A.
PD   30-DEC-1987.
PF   22-JUN-1987; E00326.
PR   23-JUN-1986; EP-401380.
PA   (INSP) Inst Pasteur.
PI   Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR   WPI; 88-014396/02.
DR   N-PSDB; N80437.
PT   New variants of lymphadenopathy associated virus (LAV) -
PT   used for prodn. of DNA, antigens and antibodies used in
PT   diagnosis of AIDS and pre-AIDS
PS   Claim 8; Fig 8A-8I; 72pp; English.
CC   LAV EL I (N80436) and LAV MA L (N80437) were isolated from the peripheral
CC   blood lymphocytes of patients. Different AIDS virus isolates concerned
CC   are designated by 3 letters of the patients name. Stable probes including
CC   the DNA sequences can be used for detection of the new LAV viruses or
CC   related viruses or DNA proviruses in eg. biological samples. The proteins
CC   or peptides can be used for detection of antibodies induced in vivo and
CC   present in biological fluids. The DNA can also be used for the expression
CC   of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC   polypeptides can also be used for the prodn. of antibodies for the
CC   detection of proteins related to the LAV viruses, partic. for diagnosis
CC   of AIDS or pre-AIDS.
SQ   Sequence   505 AA;

  Query Match           2.8%;  Score 96;  DB 1;  Length 505;
  Best Local Similarity  41.9%;  Pred. No. 1.61e+01;
  Matches   18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db     28 kyr-lkhlvwasrelerfalnpglletgegcqqimeql-qstl 68
          ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT    3
ID   R15252 standard; Protein; 32 AA.
AC   R15252;
DT   12-FEB-1992  (first entry)
DE   Carbohydrate binding domain #9.
KW   cellulose; CBD; hemicellulosic substrate;
KW   Trichoderma reesei; cellulase; terminal A region.
PN   WO9117244-A.
PD   14-NOV-1991.
PF   08-MAY-1991; DK0124.
PR   09-MAY-1990; DK-001158.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Woldike HF, Hagen F, Hjort CM, Hastrup S;
DR   WPI; 91-353766/48.
PT   New fungal (hemi)cellulose degrading enzymes - for prodn. of liq.
PT   fuel gas and feed protein, have specified carbohydrate binding domain
PS   Claim 20; Page 45; 73pp; English.
CC   This CBD is homologous to a terminal A region of Trichoderma reesei
CC   cellulases and effects binding of a protein to an insoluble
CC   cellulosic or hemicellulosic substrate. It is one of ten specific
CC   CBD's (see R15244-R15253) which correspond to the generic CBD
CC   formulae in R15242 and R15243. The CBD is incorporated into a fusion
CC   protein comprising a catalytic domain from a cellulase, e.g. a
CC   Bacillus endoglucanase, and optionally comprising a linking B domain
CC   from e.g. a fungal endoglucanase.
SQ   Sequence   32 AA;

  Query Match           2.6%;  Score 87;  DB 3;  Length 32;
  Best Local Similarity  35.0%;  Pred. No. 7.18e+01;
  Matches    7;  Conservative    8;  Mismatches   4;  Indels   1;  Gaps   1;

Db     14 tckspftck-kindfysqcq 32
          :||: : |: :: ||:: ||
Qy    249 SCKTNYICRSRLADFFTNCQ 268


RESULT    4
ID   W43305 standard; protein; 119 AA.
AC   W43305;
DT   01-APR-1998  (first entry)
DE   p17 antigen of human immunodeficiency virus type 1.
KW   Immunoassay; HIV-1; human immunodeficiency virus type 1; saliva;
KW   immobilised; p17 antigen; detection; nitrocellulose support; antibody.
OS   Human immunodeficiency virus type 1.
PN   US5695930-A.
PD   09-DEC-1997.
PF   05-AUG-1996; 692445.
PR   10-NOV-1994; US-337670.
PA   (KILP/) KILPATRICK T J.
PA   (WEIN/) WEINSTEIN D E.
PI   Kilpatrick TJ, Weinstein DE;
DR   WPI; 98-041291/04.
PT   Immunoassay for HIV-1 antibodies in saliva samples - using
PT   immobilised p17 antigen
```

```
PS   Claim 1; Columns 11-14; 12pp; English.
CC   The present sequence represents the p17 protein antigen of human
CC   immunodeficiency virus type 1 (HIV-1). The invention relates to a method
CC   for detecting antibodies to HIV-1, which comprises contacting saliva
CC   from a fasting human with immobilised HIV-1 p17 antigen, preferably on a
CC   nitrocellulose-containing solid support, and detecting any antigen-
CC   antibody complex, optionally with a labelled second antibody. The method
CC   can be used to detected exposure of a mammal, especially a human, to
CC   HIV-1. Detecting anti-HIV antibodies in saliva is safer as HIV is not
CC   transmitted via saliva, as opposed to blood. The assay is simple and
CC   cheap to perform without laboratory equipment or trained personnel and
CC   can be performed at home by the saliva donor.
SQ   Sequence   119 AA;

  Query Match             2.6%;  Score 86;  DB 27;  Length 119;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     27 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 67
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
       45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT    5
ID   P70272 standard; protein; 120 AA.
AC   P70272;
DT   27-FEB-1991  (first entry)
DE   The sequence encoding the HIV virus 14 KDa gag precursor
DE   protein.
KW   HIV virus; gag protein; AIDS; vaccine.
OS   HIV virus.
PN   EP-230222-A.
PD   29-JUL-1987.
PF   05-JAN-1987; 100064.
PR   06-JAN-1986; US-816645.
PA   (HOFF ) HOFFMANN-LA ROCHE AG.
PI   Kramer R, Reddy P, Shaber M;
DR   WPI; 87-207550/30.
DR   N-PSDB; N70431.
PT   Polypeptide obtd. from HTLV-III gag-gene - used for detecting
PT   AIDS virus and for prepn. of vaccines giving protective immunity.
PS   Claim 3; Page 13; 39pp; English.
CC   The sequence encodes the HIV virus 14 KDA gag precursor gene.
CC   The HIV virus gag protein or proteolytically derived products
CC   may be used as a vaccine against HIV virus infection (AIDS).
CC   See also N70427, N70428, N70429, N70430 and N70432.
SQ   Sequence   120 AA;

  Query Match             2.6%;  Score 86;  DB 2;  Length 120;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     16 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 56
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT    6
ID   R26326 standard; protein; 132 AA.
AC   R26326;
DT   04-FEB-1993  (first entry)
DE   HTLV-IIIb p17 sequence.
KW   human immunodeficiency virus; HIV-1; HIV-2; glycoprotein; gag;
KW   epitope mapping; monoclonal antibody; G11c1; G11h3; B4F8; D3b3.
OS   Human t-cell lymphotropic virus.
FH   Key             Location/Qualifiers
FT   peptide         1..15
FT                   /label= #1
FT   peptide         11..25
FT                   /label= #2
FT   peptide         21..35
FT                   /label= #3
FT   peptide         31..46
FT                   /label= #4
FT   peptide         41..55
FT                   /label= #5
FT   peptide         51..65
FT                   /label= #6
FT   peptide         62..78
FT                   /label= #7
FT   peptide         73..86
FT                   /label= #8
FT   peptide         78..92
FT                   /label= #9
FT   peptide         86..115
FT                   /label= #9A
FT   peptide         93..107
FT                   /label= #10
FT   peptide         105..115
FT                   /label= #11
FT   peptide         111..123
FT                   /label= #12
FT   peptide         121..132
FT                   /label= #13
PN   WO9212731-A.
PD   06-AUG-1992.
PF   22-JAN-1992; U00528.
PR   22-JAN-1991; US-644435.
PA   (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
PI   Abraham P, Honnen WJ, Revesz K, Shang F;
DR   WPI; 92-284421/34.
PT   Pure glycoprotein contg. gag amino acid sequences - expressed on
PT   surface of HIV-infected cells and monoclonal antibodies to the
PT   glycoprotein,for immunoassay and protecting against HIV infection
PS   Claim 14; Fig 12; 51pp; English.
CC   The synthetic peptides #1-13 were used in epitope mapping studies
CC   of monoclonal antibodies against p17. ELISA plates were coated with
CC   recombinant p17 (rp17) and peptides #1-13 (representing the complete
CC   sequence of HTLV-IIIb p17). Antibodies G11g1 and G11h3 (which react
CC   with live HIV infected cells and recognise a glycoprotein contg. a
CC   gag amino acid sequence and having mol. wt. 150,000 or 90,000,
CC   respectively), bound strongly to rp17 but did not bind to any of the
CC   peptides tested. The claimed monoclonal antibodies which competitively
CC   inhibit G11g1 or G11h3 are further defined by the fact that they do
CC   not substantially bind to the peptides 1-13. The MAbs can be used
CC   to determine the number of HIV-infected cells, to purify gag
CC   antigens or can be coupled to a toxin for therapeutic use.
SQ   Sequence   132 AA;

  Query Match             2.6%;  Score 86;  DB 5;  Length 132;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 68
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT    7
ID   R55851 standard; Protein; 169 AA.
AC   R55851;
DT   07-DEC-1994  (first entry)
DE   HIV-1 MA delta-5-16 mutant.
KW   HIV-1; matrix protein; MA; gene therapy; AIDS; deletion mutant.
OS   Human immunodeficiency virus type 1.
PN   WO9412513-A.
PD   09-JUN-1994.
PF   12-JAN-1993; U00267.
PR   23-NOV-1992; US-979966.
PA   (HARD ) HARVARD COLLEGE.
PI   Essex ME,  Lee TH,  Yu X;
DR   WPI; 94-200178/24.
PT   Treating HIV infection with mutant HIV matrix polypeptide - pref.
PT   produced by transfection with viral vector, inhibiting
PT   replication, assembly or infectivity of virus particles
PS   Disclosure; Fig. 1A-1F; 61pp; English.
```

```
CC   HIV infection is treated by administering to a patient a mutated
CC   HIV matrix (MA) protein using a viral vector.  Preferred mutants
CC   are those given in R55851-60.  The deletion/substitutions of the
CC   mutants either lower the incorporation of Env protein in the viral
CC   particle in HIV infected cells, disrupt virus assembly, or disrupt
CC   viral entry into uninfected cells.  The DNA sequence encoding MA
CC   is given in Q66839 and the aa sequence in R55850.
SQ   Sequence   169 AA;

  Query Match           2.6%;  Score 86;  DB 10;  Length 169;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches   16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     16 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 56
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


    LT    8
     R55857 standard; Protein; 171 AA.
     R55857;
DT   07-DEC-1994  (first entry)
DE   HIV-1 MA delta-105-114 mutant.
KW   HIV-1; matrix protein; MA; gene therapy; AIDS; deletion mutant.
OS   Human immunodeficiency virus type 1.
PN   WO9412513-A.
PD   09-JUN-1994.
PF   12-JAN-1993; U00267.
PR   23-NOV-1992; US-979966.
PA   (HARD ) HARVARD COLLEGE.
PI   Essex ME,  Lee TH,  Yu X;
DR   WPI; 94-200178/24.
PT   Treating HIV infection with mutant HIV matrix polypeptide - pref.
PT   produced by transfection with viral vector, inhibiting
PT   replication, assembly or infectivity of virus particles
PS   Disclosure; Fig. 1A-1F; 61pp; English.
CC   HIV infection is treated by administering to a patient a mutated
CC   HIV matrix (MA) protein using a viral vector.  Preferred mutants
CC   are those given in R55851-60.  The deletion/substitutions of the
CC   mutants either lower the incorporation of Env protein in the viral
CC   particle in HIV infected cells, disrupt virus assembly, or disrupt
CC   viral entry into uninfected cells.  The DNA sequence encoding MA
CC   is given in Q66839 and the aa sequence in R55850.
SQ   Sequence   171 AA;

  Query Match           2.6%;  Score 86;  DB 10;  Length 171;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches   16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 68
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT    9
ID   P94509 standard; protein; 267 AA.
AC   P94509;
DT   10-MAR-1993  (revised)
DT   19-JUN-1989  (first entry)
DE   Sequence encoded by plasmid pUC18Xmnstop in E.coli cells.
KW   Human immunodeficiency virus; AIDS; bacterial lysate; HIV antigen.
FH   Key             Location/Qualifiers
FT   cds             1..800
FT   misc_feature    51..484
FT                   /label=p17 coding region
PN   DE3724016-A.
PD   02-FEB-1989.
PF   21-JUL-1987; 724016.
PR   21-JUL-1987; DE-724016.
PA   (WOLF) Wolf H.
PI   Soutschek-Bauer E, Motz M, Wolf H;
DR   WPI; 89-039954/06.
DR   N-PSDB; N91141.
PT   New DNA contg. HIV coding sequences used as source of antigen for Western
PT   blot diagnosis for AIDS, usable directly as bacterial lysate.
PS   Disclosure; fig.7; 40pp; German.
CC   This sequence is that of the human immunodeficiency virus-1 (HIV-1) p17
CC   antigen which is expressed as a fusion protein with a few extra AAs
CC   on both ends. The antigen (Ag) can be used in a western blot analysis to
CC   detect HIV antibodies in serum or other body fluids. The prepn. is both
CC   cheap and safe and the Ag is used directly as a bacterial lysate. The
CC   diagnosis is more sensitive and accurate than conventional western blot
CC   methods.
SQ   Sequence   267 AA;

  Query Match           2.6%;  Score 86;  DB 2;  Length 267;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches   16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     27 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 67
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   10
ID   W19681 standard; Protein; 446 AA.
AC   W19681;
DT   20-AUG-1997  (first entry)
DE   ATM mutant C1339T.
KW   ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW   central nervous system; immune system; chromosomal instability; therapy;
KW   cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW   multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW   general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW   blood vessel; bulbar conjunctiva; facial skin; A-T; mutein.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_difference 263
FT                   /note= "encoded by CTT"
PN   WO9636691-A1.
PD   21-NOV-1996.
PF   16-MAY-1996; U07025.
PR   16-MAY-1995; US-441822.
PR   08-APR-1996; US-629001.
PA   (KOHN/) KOHN K I.
PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI   Shiloh Y;
DR   WPI; 97-012070/01.
DR   N-PSDB; T68764.
PT   New isolated ataxia-telangiectasia gene - used to develop prods. for
PT   the study, diagnosis and treatment of ataxia-telangiectasia.
PS   Claim 5; Page -; 153pp; English.
CC   W19648-W19702 represent mutations of the ATM gene of the invention (see
CC   W06234 for wild type sequence). The ATM gene of the invention, is located
CC   in the human genome at chromosome segment 11q22-23. Ataxia-telangiectasia
CC   (A-T) is caused by mutations, insertions, or deletions (such as
CC   represented by this sequence) in the coding region of the ATM gene. A-T
CC   is a progressive genetic disorder affecting the central nervous and
CC   immune systems. A-T involves chromosomal instability, cancer
CC   predisposition, radiation sensitivity, and cell cycle abnormalities. A-T
CC   is a multi-system disease inherited in an autosomal recessive manner. The
CC   wild type gene, can be used in methods for detecting carriers of a
CC   defective gene that causes A-T. The gene can also be used to generate
CC   antibodies. The methods and antibodies can be used in the study,
CC   diagnosis and therapy of A-T.
SQ   Sequence   446 AA;

  Query Match           2.6%;  Score 86;  DB 22;  Length 446;
  Best Local Similarity  31.1%;  Pred. No. 8.45e+01;
  Matches   14;  Conservative   14;  Mismatches  15;  Indels   2;  Gaps   2;

Db    399 hlqksqndfdlvpwlqiatqliskypaslpncelspll-milsql 442
          ||  |  :::    |  :::: :|  |: |:| ||||| :::: |
Qy    412 HLCISNGNYEKEG-LGASSHITTKSMAAPPSCGLSPLLVLVVTAL 455
```

```
RESULT   11
ID   P60309 standard; Protein; 478 AA.
AC   P60309;
DT   08-JUN-1991  (first entry)
DE   Sequence of gag and p24gag.
KW   HIV; LAV; AIDS; diagnosis; vaccine.
OS   HTLV-IIIB/H9 cells (ATCC CRL 8543).
FH   Key             Location/Qualifiers
FT   protein         1..132
FT                   /label= gag
FT   protein         133..478
FT                   /label= p24gag
PN   EP-187041-A.
PD   09-JUL-1986.
PF   23-DEC-1985; 309454.
PR   24-DEC-1984; US-685272.
PR   04-DEC-1985; US-805069.
PA   (GETH ) GENENTECH INC.
PI   Capon DJ, Lasky LA;
DR   WPI; 86-177602/28.
DR   N-PSDB; N60288
PT   Acquired immune deficiency syndrome polypeptide(s) - obtd. by
PT   molecular cloning etc. and used for diagnosis and in vaccines
PT   against virus disease
PS   Example; fig 2; 125pp; English.
CC   A comparison of N60287 with the cDNA of the HTLV-III genome
CC   revealed one particular clone, designated p7.11 which contained a
CC   DNA sequence encoding this peptide (P60308) sequence.  This approx.
CC   2.2 kilobase covers the precursor gag region and encodes, 5' to 3',
CC   p-12, p-15, p-24 a second p-15 protein, and approx. 300 extra base
CC   pairs 3' to the gag region (see N60288).
SQ   Sequence   478 AA;

  Query Match             2.6%;  Score 86;  DB 3;  Length 478;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 68
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   12
ID   P93707 standard; Protein; 504 AA.
AC   P93707;
DT   11-MAR-1992  (first entry)
DE   Sequence of the gag protein precursor.
KW   Vaccine; diagnosis; AIDS; LAV; HTLV-III.
     Human immunodeficiency virus.
     EP-345242-A.
     06-DEC-1989.
PF   02-JUN-1989; 870082.
PR   03-JUN-1988; US-202271.
PR   30-JUN-1989; AU-037256.
PR   11-JUL-1989; WO-U02415.
PA   (SMIK ) SMITHKLINE BIOLOGIC.
PI   Gheysen DA, Jacobs E;
DR   WPI; 89-358596/49.
DR   N-PSDB; N92575.
PT   New recombinant DNA encoding HIV gag precursor protein - without
PT   flanking sequences, expressed in insect, yeast and mammalian
PT   cells, producing particles useful in vaccines and diagnosis
PS   Disclosure; Page 4-6; 17pp; French.
CC   The inventors claim recombinant gag protein precursor (P93707) and
CC   the DNA encoding it (N92575) operatively linked to regulatory
CC   elements functional in eukaryotic cells.
SQ   Sequence   504 AA;

  Query Match             2.6%;  Score 87;  DB 3;  Length 504;
  Best Local Similarity  37.2%;  Pred. No. 7.18e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 kyk-lkhivwasrelerfavnpglletsegcreilgql-qpsl 68
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   13
ID   R29704 standard; Protein; 506 AA.
AC   R29704;
DT   26-APR-1993  (first entry)
DE   gag gene decoded from viral DNA from hTLR.
KW   express proteins; lymphadenopathy syndrome; AIDS; HIV; HTLV;
KW   serological immunoassays; antibodies to hTLR; monoclonal antibodies;
KW   probes; ss.
OS   Human T cell lymphotropic retrovirus.
PN   EP-518443-A.
PD   16-DEC-1992.
PF   30-OCT-1985; 201711.
PR   31-OCT-1984; US-667501.
PR   30-JAN-1985; US-696534.
PR   06-SEP-1985; US-773447.
PA   (CHIR ) CHIRON CORP.
PI   Barr PJ, Dina D, George-Nascimento C, Hallewell R;
PI   Luciw PA, Parkes D, Pescador RS, Steimer K, Truett M;
DR   WPI; 92-417329/51.
DR   N-PSDB; Q31936.
PT   Recombinant DNA construct including replication system recognised
PT   by unicellular microorganism - used to form recombinant proteins
PT   for diagnosing AIDS and lymphadenopathy syndrome
PS   Example 5; Fig 2; 32pp; English.
CC   This sequence was decoded from the gag gene from proviral hTLR DNA.
CC   Proteins associated with lymphadenopathy syndrome and/or AIDS may
CC   be used in serological immuno-assays to detect antibodies to hTLR.
CC   The polypeptides can be used alone or in fusion constructs to
CC   produce antisera or monoclonal antibodies which may be used for
CC   therapy or diagnosis.
SQ   Sequence   506 AA;

  Query Match             2.6%;  Score 86;  DB 6;  Length 506;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     32 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 72
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   14
ID   R10866 standard; Protein; 512 AA.
AC   R10866;
DT   17-APR-1991  (first entry)
DE   HIV gag precursor protein.
KW   human immunodeficiency virus; HIV; gag protein; AIDS vaccine.
PN   WO9100904-A.
PN   J03049686-A (non-conventional equivalent).
PD   24-JAN-1991.
PF   11-JUL-1989; U02415.
PR   03-JUN-1988; US-202271.
PR   11-JUL-1989; WO-U02415.
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PI   Gheysen D, Jacobs E;
DR   WPI; 91-051332/07.
DR   N-PSDB; Q10574.
PT   DNA encoding HIV gag precursor - devoid of natural flanking
PT   regions and with regulatory, regions for expression in eukaryotic
PT   cells, used for vaccine prepn.
PS   Disclosure; Page 7; 40pp; English.
CC   The protein is encoded by an exemplary coding region for the HIV
CC   gag precursor which lacks 3' and 5' flanking sequences. For
CC   expression in eucaryotic cells, the sequence is operatively linked
CC   to a regulatory element which functions in the selected host cells.
CC   Immunogenic particles comprising full-length gag precursor protein
CC   are immunologically similar to authentic gag particles. The protein
CC   is recognised by anti-p16, anti-p24 and anti-p17 antibodies and
CC   lacks functions required for viral maturation and replication.
```

```
CC   It can be used as a diagnostic agent for detection of exposure to
CC   HIV.
SQ   Sequence   512 AA;

  Query Match            2.6%;  Score 86;  DB 2;  Length 512;
  Best Local Similarity  37.2%;  Pred. No. 8.45e+01;
  Matches   16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 kyk-lkhivwasrelerfavnpglletsegcrqilgql-qpsl 68
          ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   15
ID   R15237 standard; Protein; 514 AA.
AC   R15237;
DT   11-FEB-1992  (first entry)
DE   Fusarium oxysporum C-family cellobiohydrolase.
KW   cellulose; carbohydrate binding domain; CBD; fungi; cellulase.
OS   Fusarium oxysporum.
FH   Key             Location/Qualifiers
FT   domain          483..514
FT                   /label= CBD
FT                   /note= "homologous to a terminal A region of
FT                   T.reesei cellulase"
PN   WO9117244-A.
PD   14-NOV-1991.
PF   08-MAY-1991; DK0124.
PR   09-MAY-1990; DK-001158.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Woldike HF, Hagen F, Hjort CM, Hastrup S;
DR   WPI; 91-353766/48.
DR   N-PSDB; Q14838.
PT   New fungal (hemi) cellulose degrading enzymes - for prodn. of
PT   liq. fuel gas and feed protein, have specified carbohydrate
PT   binding domain
PS   Claim 1; Fig 11; 73pp; English.
CC   The full-length C-family cellobiohydrolase coding sequence was
CC   isolated from a F.oxysporum cDNA library by hybridisation to PCR
CC   generated genomic oligonucleotide probes (see esp. Q15502). The
CC   enzyme encoded by the sequence has a CBD which is homologous to a
CC   terminal A region of T.reesei cellulase and is able to bind to
CC   insoluble (hemi)cellulosic substrates. See also Q14839-Q14842.
SQ   Sequence   514 AA;

  Query Match            2.6%;  Score 87;  DB 3;  Length 514;
  Best Local Similarity  35.0%;  Pred. No. 7.18e+01;
  Matches    7;  Conservative    8;  Mismatches   4;  Indels   1;  Gaps   1;

Db    496 tckspftck-kindfysqcq 514
          :||: : |: :: ||:: ||
Qy    249 SCKTNYICRSRLADFFTNCQ 268


Search completed: Tue Jun 23 18:28:25 1998
Job time : 146 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted] on: Tue Jun 23 18:23:39 1998; MasPar time 23.53 Seconds
721.817 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-866-354-2
Description: (1-465) from US08866354.pep
Perfect Score: 3369
Sequence: 1 MFLATLYFALPLLDLLLSAE.....PLLVLVVTALSTLLSLTETS 465

Scoring table: PAM 150
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir56
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 46.109; Variance 80.746; scale 0.571

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

[redacted]sult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	171	5.1	24	2	S69080	glycosyl-phosphatidyl	3.50e-13
2	102	3.0	249	2	S09868	hypothetical protein	1.21e-01
3	101	3.0	352	2	S60024	bradykinin B1 recepto	1.68e-01
4	98	2.9	346	2	A70144	hypothetical protein	4.43e-01
5	93	2.8	169	2	I64089	protein-export protei	2.13e+00
6	95	2.8	336	2	D70030	alkanal monooxygenase	1.15e+00
7	94	2.8	411	2	H69158	LPS biosynthesis RfbU	1.56e+00
8	93	2.8	600	2	S07638	spore coat protein SP	2.13e+00
9	95	2.8	706	2	S62933	hypothetical protein	1.15e+00
10	96	2.8	826	2	A60385	monocyte surface anti	8.37e-01
11	95	2.8	857	2	S33821	median body protein -	1.15e+00
12	94	2.8	1075	2	S54067	probable membrane pro	1.56e+00
13	91	2.7	475	2	S49886	probable membrane pro	3.92e+00
14	92	2.7	493	2	JC5621	epidermal growth fact	2.89e+00
15	92	2.7	563	1	VCMVM7	env polyprotein - bab	2.89e+00
16	92	2.7	688	1	C1HUS	complement subcompone	2.89e+00
17	92	2.7	695	2	S05008	serine proteinase, ca	2.89e+00
18	86	2.6	120	5	1TAM	hiv-1 matrix protein	1.71e+01
19	86	2.6	167	2	C64361	hypothetical protein	1.71e+01
20	86	2.6	316	2	S58719	probable membrane pro	1.71e+01
21	89	2.6	327	2	A55356	urokinase-type plasmi	7.12e+00
22	87	2.6	342	2	S42885	beta-1,3-glucanase (E	1.28e+01
23	86	2.6	387	2	B49175	Motch A protein - mou	1.71e+01
24	86	2.6	478	1	FOVWVL	gag polyprotein - hum	1.71e+01
25	89	2.6	491	2	S49779	CDC1 protein - yeast	7.12e+00
26	88	2.6	494	2	G64382	acetolactate synthase	9.56e+00
27	86	2.6	500	1	FOVWLV	gag polyprotein - hum	1.71e+01
28	86	2.6	502	1	FOVWA2	gag polyprotein - hum	1.71e+01
29	88	2.6	506	1	A38068	gag polyprotein - hum	9.56e+00
30	86	2.6	512	1	FOVWH3	gag polyprotein - hum	1.71e+01
31	88	2.6	521	1	FOLJST	gag polyprotein - hum	9.56e+00
32	88	2.6	521	2	S53091	gag polyprotein - hum	9.56e+00
33	87	2.6	521	2	S12152	gag polyprotein - hum	1.28e+01
34	88	2.6	522	1	FOLJGG	gag polyprotein - hum	9.56e+00
35	87	2.6	642	2	G69371	acetyl-CoA synthetase	1.28e+01
36	86	2.6	646	2	S36586	E1 protein - human pa	1.71e+01
37	86	2.6	769	1	JC1121	leukocyte adhesion pr	1.71e+01
38	89	2.6	1033	2	S54506	probable membrane pro	7.12e+00
39	87	2.6	1146	2	B35962	protein-tyrosine kina	1.28e+01
40	87	2.6	1182	2	A35962	protein-tyrosine kina	1.28e+01
41	88	2.6	1403	2	S24548	homeotic protein pros	9.56e+00
42	88	2.6	1403	2	JQ1397	pros protein - fruit	9.56e+00
43	88	2.6	1407	2	A41089	neuronal precursor pr	9.56e+00
44	87	2.6	2471	2	A49128	cell-fate determining	1.28e+01
45	87	2.6	2531	2	S18188	notch protein homolog	1.28e+01

ALIGNMENTS

RESULT 1
ENTRY S69080 #type complete
TITLE glycosyl-phosphatidylinositol-linked protein GDNFR-alpha - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 13-Mar-1998
ACCESSIONS S69080
REFERENCE S69080
#authors Treanor, J.J.S.; Goodman, L.; de Sauvage, F.; Stone, D.M.; Poulsen, K.T.; Beck, C.D.; Gray, C.; Armanini, M.P.; Pollock, R.A.; Hefti, F.; Phillips, H.S.; Goddard, A.; Moore, M.W.; Buj-Bello, A.; Davies, A.M.; Asai, N.; Takahashi, M.; Vandlen, R.; Henderson, C.E.; Rosenthal, A.
#journal Nature (1996) 382:80-83
#title Characterization of a multicomponent receptor for GDNF.
#accession S69080
##status preliminary
##molecule_type mRNA
##residues 1-24 ##label TRE
SUMMARY #length 24 #molecular-weight 2573 #checksum 2584

Query Match 5.1%; Score 171; DB 2; Length 24;
Best Local Similarity 95.8%; Pred. No. 3.50e-13;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 MFLATLYFALPLLDLLMSAEVSGG 24
||||||||||||||||:|||||||
Qy 1 MFLATLYFALPLLDLLLSAEVSGG 24

RESULT 2
ENTRY S09868 #type complete
TITLE hypothetical protein UL103 - human cytomegalovirus (strain AD169)
ORGANISM #formal_name human cytomegalovirus, human herpesvirus 5
#note host Homo sapiens (man)
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Sep-1997
ACCESSIONS S09868
REFERENCE S09749
#authors Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; Hutchison III, C.A.; Kouzarides, T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.; Tomlinson, P.; Weston, K.M.; Barrell, B.G.
#journal Curr. Top. Microbiol. Immunol. (1990) 154:125-169
#title Analysis of the protein-coding content of the sequence of

```
                human cytomegalovirus strain AD169.
   #cross-references MUID:90269039
   #accession     S09868
      ##status          nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues        1-249 ##label CHE
      ##cross-references EMBL:X17403; NID:g59591; PID:g1780882
      ##note            this sequence was submitted to the EMBL Data Library,
                         December 1989
SUMMARY           #length 249  #molecular-weight 28636  #checksum 1456

  Query Match             3.0%;  Score 102;  DB 2;  Length 249;
  Best Local Similarity  36.2%;  Pred. No. 1.21e-01;
  Matches    17;  Conservative    8;  Mismatches  20;  Indels   2;  Gaps   2;

Db    106 CTVISCVENCNLTRKCLHDLLQYLDAVNVRESFGRLLHHSARRLICS 152
          ||     : ||  |||   | |::| |  : |:| :|  | | : |:
Qy    178 CTTSVSNDVCN-RRKCHKALRQFFDKVPAKHSYG-MLFCSCRDIACT 222


RESULT    3
ENTRY             S60024     #type complete
TITLE             bradykinin B1 receptor - rabbit
ORGANISM          #formal_name Oryctolagus cuniculus #common_name domestic
                    rabbit
DATE              24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change
                    10-Sep-1997
ACCESSIONS        S60024
REFERENCE         S60024
   #authors       MacNeil, T.; Bierilo, K.K.; Menke, J.G.; Hess, J.F.
   #journal       Biochim. Biophys. Acta (1995) 1264:223-228
   #title         Cloning and pharmacological characterization of a rabbit
                    bradykinin B(1) receptor.
   #accession     S60024
      ##status          preliminary
      ##molecule_type mRNA
      ##residues        1-352 ##label MAC
      ##cross-references EMBL:U20507; NID:g1041820; PID:g1041821
KEYWORDS          G protein-coupled receptor; transmembrane protein
SUMMARY           #length 352  #molecular-weight 39503  #checksum 8863

  Query Match             3.0%;  Score 101;  DB 2;  Length 352;
  Best Local Similarity  26.0%;  Pred. No. 1.68e-01;
  Matches    26;  Conservative   32;  Mismatches  32;  Indels  10;  Gaps  10;

Db    216 AILFFNCHILA-SLRRRGERVPSRCG-GPRDSKSTALIL-TLVASFLVCWAPYHFFAFLE 272
          ::|| :|: :| : |||   ||  |:   |: |:  | |     :  :|::    | | :
      210 GMLFCSCRDIACTERRRQTIVPV-CSYEERE-KPNCLNLQDSCKTNYICRSRLADF-FTN 266

      273 CLWQVHAIGGCFWEEFTDLGLQLSNFSAFVNSCLNPVIYV 312
          |  : :::::|: |:::|  | |: :|::: : : |  |:
Qy    267 CQPESRSVSSCLKENYAD-CL-LA-YSGLIGTVMTPN-YI 302


RESULT    4
ENTRY             A70144     #type complete
TITLE             hypothetical protein BB0354 - Lyme disease spirochete
ORGANISM          #formal_name Borrelia burgdorferi #common_name Lyme disease
                    spirochete
DATE              13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                    13-Feb-1998
ACCESSIONS        A70144
REFERENCE         A70100
   #authors       Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
                    Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
                    Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
                    J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
                    Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
                    Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
                    J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.;
                    Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
                    K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
   #journal       Nature (1997) 390:580-586
   #title         Genomic sequence of a Lyme disease spirochaete, Borrelia
                    burgdorferi.
   #cross-references MUID:98065943
   #accession     A70144
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues        1-346 ##label KLE
      ##cross-references GB:AE000783; TIGR:BB0354
      ##experimental_source strain B31
SUMMARY           #length 346  #molecular-weight 41137  #checksum 5498

  Query Match             2.9%;  Score 98;  DB 2;  Length 346;
  Best Local Similarity  31.9%;  Pred. No. 4.43e-01;
  Matches    15;  Conservative   11;  Mismatches  19;  Indels   2;  Gaps   2;

Db      1 MKRLKRRLKIFWSITKRLKVFSIFEGGGF-AVFDSLRLIFLIICRFI 46
            ||: |  |:|:||:   |    ::| : : :|   |  || ::  ||
Qy     93 MKKEKNCLRIYWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVV-PFI 138


RESULT    5
ENTRY             I64089     #type complete
TITLE             protein-export protein - Haemophilus influenzae (strain Rd
                    KW20)
ORGANISM          #formal_name Haemophilus influenzae
DATE              18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
                    24-Oct-1997
ACCESSIONS        I64089
REFERENCE         A64000
   #authors       Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
                    Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
                    Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
                    FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
                    Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
                    J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
                    M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
                    D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
                    J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
                    Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
   #journal       Science (1995) 269:496-512
   #title         Whole-genome random sequencing and assembly of Haemophilus
                    influenzae Rd.
   #cross-references MUID:95350630
   #accession     I64089
      ##status          nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues        1-169 ##label TIGR
      ##cross-references GB:U32758; GB:L42023; NID:g1573747; PID:g1573750;
                          TIGR:HI0743
FUNCTION
   #description   one of the proteins required for the normal export of
                    envelope proteins out of the cell cytoplasm; may be
                    involved in the initiation of the exporting process
CLASSIFICATION    #superfamily protein-export protein secB
KEYWORDS          protein export
SUMMARY           #length 169  #molecular-weight 19132  #checksum 8441

  Query Match             2.8%;  Score 93;  DB 2;  Length 169;
  Best Local Similarity  25.0%;  Pred. No. 2.13e+00;
  Matches    16;  Conservative   17;  Mismatches  30;  Indels   1;  Gaps   1;

Db     75 LEDSGDVAFICEVKQAGVFTISGLEDVQMAHCLTSQCPNMLFPYARELVSNLVNRGTFPA 134
          |:||    :||  : |  ||    |   :: ||    :: |::|:  |:: ::    : :
Qy    246 LQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYS-GLIGTVMTPNYIDS 304

Db    135 LNLS 138
           :||
Qy    305 SSLS 308


RESULT    6
ENTRY             D70030     #type complete
```

```
TITLE           alkanal monooxygenase homolog yvbT - Bacillus subtilis
ORGANISM        #formal_name Bacillus subtilis
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                  26-Feb-1998
ACCESSIONS      D70030
REFERENCE       A69580
  #authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                  Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                  Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                  A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                  Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                  Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                  Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                  Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                  Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                  M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                  S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                  Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                  C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                  Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                  Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                  Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                  Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                  Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                  Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                  M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                  M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                  V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                  A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                  Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                  Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                  Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                  Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                  B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                  Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                  Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                  Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                  Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                  Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                  K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                  Yoshikawa, H.; Danchin, A.
  #journal      Nature (1997) 390:249-256
  #title        The complete genome sequence of the Gram-positive bacterium
                  Bacillus subtilis.
  #accession    D70030
    ##status         preliminary; nucleic acid sequence not shown;
                      translation not shown
    ##molecule_type DNA
    ##residues      1-336 ##label KUN
    ##experimental_source strain 168
GENETICS
  #gene         yvbT
CLASSIFICATION  #superfamily ynbW protein
SUMMARY         #length 336  #molecular-weight 37101  #checksum 2405

  Query Match             2.8%;  Score 95;  DB 2;  Length 336;
  Best Local Similarity  42.5%;  Pred. No. 1.15e+00;
  Matches    17;  Conservative   11;  Mismatches  10;  Indels   2;  Gaps   2;

Db    132 NSGEDFPEQLEELRNYFKPSGNVRNQVRAIPGEGIDVPIW 171
          |||:|: | |  | |:|| :  ::| ::|: |:| || :|
Qy    317 NSGNDLEECLKFL-NFFKDNTCLKNAIQAF-GNGSDVTVW 354


RESULT    7
ENTRY           H69158     #type complete
TITLE           LPS biosynthesis RfbU related protein - Methanobacterium
                  thermoautotrophicum (strain Delta H)
ORGANISM        #formal_name Methanobacterium thermoautotrophicum
DATE            05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                  05-Dec-1997
ACCESSIONS      H69158
REFERENCE       A69000
  #authors      Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
                  Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
                  Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
                  Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
                  Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
                  A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
                  McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
                  Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
                  J.; Reeve, J.N.
  #journal      J. Bacteriol. (1997) 179:7135-7155
  #title        Complete genome sequence of Methanobacterium
                  thermoautotrophicum Delta H: functional analysis and
                  comparative genomics.
  #cross-references MUID:98037514
  #accession    H69158
    ##status         preliminary; nucleic acid sequence not shown;
                      translation not shown
    ##molecule_type DNA
    ##residues      1-411 ##label MTH
    ##cross-references GB:AE000666
    ##experimental_source strain Delta H
GENETICS
  #gene         MTH450
SUMMARY         #length 411  #molecular-weight 45966  #checksum 9796

  Query Match             2.8%;  Score 94;  DB 2;  Length 411;
  Best Local Similarity  24.3%;  Pred. No. 1.56e+00;
  Matches    18;  Conservative   26;  Mismatches  25;  Indels   5;  Gaps   4;

Db    252 IESLKDLELDDVKCIILGAPVDEEYLRTLRGLAESMDSDVEFFGPVPYTEVPTWMNAADV 311
          :|:||:  | : :|   |   : : ||  ::: :|:::: :::   ||  | :::  :|:
Qy     76 MEALKQKSLYNCRCKR-GMKKEKNCLRIYWSMYQSLQGN-DLLEDSPYEPVNSRL--SDI 131

Db    312 F-VLPSLEEGFGLV 324
          | |:| : : |  |
Qy    132 FRVVPFISDVFQQV 145


RESULT    8
ENTRY           S07638     #type complete
TITLE           spore coat protein SP96 precursor - slime mold (Dictyostelium
                  discoideum)
ORGANISM        #formal_name Dictyostelium discoideum
DATE            07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
                  09-Sep-1997
ACCESSIONS      S07638; A60942; B60942
REFERENCE       S07638
  #authors      Fosnaugh, K.L.; Loomis, W.F.
  #journal      Nucleic Acids Res. (1989) 17:9489
  #title        Sequence of the Dictyostelium discoideum spore coat gene
                  SP96.
  #cross-references MUID:90067962
  #accession    S07638
    ##molecule_type DNA
    ##residues      1-600 ##label FOS
    ##cross-references EMBL:X16491; NID:g7373; PID:g295736
REFERENCE       A60942
  #authors      Tasaka, M.; Hasegawa, M.; Ozaki, T.; Iwabuchi, M.; Takeuchi,
                  I.
  #journal      Cell Differ. Dev. (1990) 31:1-9
  #title        Isolation and characterization of spore coat protein (sp96)
                  gene of Dictyostelium discoideum.
  #accession    A60942
    ##molecule_type DNA
    ##residues      1-155,'T',157-414,'C',416-600 ##label TAS
  #accession    B60942
    ##status         not compared with conceptual translation
    ##molecule_type mRNA
    ##residues      1-155,'T',157-414,'C',416-600 ##label TA2
GENETICS
  #introns      22/1
KEYWORDS        duplication; glycoprotein
FEATURE
```

```
   1-18              #domain signal sequence #status predicted #label SIG\
   19-600            #product spore coat protein SP96 #status predicted
                       #label MAT\
   142-185           #region 15-residue repeats\
   447-533           #region 9-residue repeats\
   568-600           #region 32-residue repeats\
   100               #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 600  #molecular-weight 59589  #checksum 1264

  Query Match             2.8%;  Score 93;  DB 2;  Length 600;
  Best Local Similarity  40.4%;  Pred. No. 2.13e+00;
  Matches    19;  Conservative    8;  Mismatches  14;  Indels   6;  Gaps   6;

Db    298 CRNIQCPTGYRCEDHNRNPICVLEERENPDNCLTCNDVNCEASGLVC 344
          ||:| | |  |   :   |:|  ||||:| |||   | :| ::  :|
Qy    216 CRDIAC-TE-R-RRQTIVPVCSYEEREKP-NCLNLQD-SCKTN-YIC 256


RESULT    9
ENTRY             S62933     #type complete
TITLE             hypothetical protein YNL021w - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   hypothetical protein N2819
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              27-Apr-1996 #sequence_revision 03-May-1996 #text_change
                    06-Feb-1998
ACCESSIONS        S62933
REFERENCE         S62920
   #authors       Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers,
                    S.
   #submission    submitted to the Protein Sequence Database, April 1996
   #accession     S62933
      ##molecule_type DNA
      ##residues      1-706 ##label AND
      ##cross-references EMBL:Z71297; NID:g1301850; PID:e239664; PID:g1301851;
                        MIPS:YNL021w
      ##experimental_source strain S288C
GENETICS
   #gene          SGD:HDA1
      ##cross-references SGD:S0004966; MIPS:YNL021w
   #map_position 14L
SUMMARY           #length 706  #molecular-weight 80069  #checksum 2534

  Query Match             2.8%;  Score 95;  DB 2;  Length 706;
  Best Local Similarity  26.6%;  Pred. No. 1.15e+00;
  Matches    21;  Conservative   23;  Mismatches  31;  Indels   4;  Gaps   3;

      330 LVIISSGFDAADGDTIGQCHVTPSCYGHMTHMLKSLARGN--LCVVLEGGYNLDAIARSA 387
          | : :  : :|  ::     || |:|    :::|||   ||  ||:  :| |: :::: |:
Qy    371 LRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS-NGNYEKEGLGASS 429

Db    388 LSVAKVLIGEPPDELPDPL 406
          :| : : |:  |: ||
Qy    430 HITTKSMAAPPSCGLS-PL 447


RESULT   10
ENTRY             A60385     #type complete
TITLE             monocyte surface antigen MS2 precursor - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change
                    13-Mar-1998
ACCESSIONS        A60385
REFERENCE         A60385
   #authors       Yoshida, S.; Setoguchi, M.; Higuchi, Y.; Akizuki, S.;
                    Yamamoto, S.
   #journal       Int. Immunol. (1990) 2:585-591
   #title         Molecular cloning of cDNA encoding MS2 antigen, a novel cell
                    surface antigen strongly expressed in murine monocytic
                    lineage.
   #accession     A60385
      ##molecule_type mRNA
      ##residues      1-826 ##label YOS
      ##cross-references EMBL:X13335
CLASSIFICATION    #superfamily disintegrin homology
KEYWORDS          glycoprotein; surface antigen; transmembrane protein
FEATURE
   1-14              #domain signal sequence #status predicted #label SIG\
   402-484           #domain disintegrin homology #label DIS\
   659-683           #domain transmembrane #status predicted #label TMM\
   330               #active_site Glu #status predicted
SUMMARY          #length 826  #molecular-weight 89896  #checksum 2686

  Query Match             2.8%;  Score 96;  DB 2;  Length 826;
  Best Local Similarity  33.3%;  Pred. No. 8.37e-01;
  Matches    17;  Conservative   10;  Mismatches  18;  Indels   6;  Gaps   6;

Db    585 ELVLQGTKC-EEGKVC-MDGSCQDLR-VYRSENCSAKCNNHGVCNHKRECH 632
          | :  | :| : :| | :|  |   |  | :  |::  :|  ||| :| ||
Qy    146 EHIPKGNNCLDAAKACNLDDICKKYRSAYITP-CTTSVSND-VCN-RRKCH 193


RESULT   11
ENTRY             S33821     #type complete
TITLE             median body protein - Giardia lamblia
ORGANISM          #formal_name Giardia lamblia
DATE              06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                    09-Sep-1997
ACCESSIONS        S33821
REFERENCE         S33821
   #authors       Marshall, J.; Holberton, D.V.
   #journal       J. Mol. Biol. (1993) 231:521-530
   #title         Sequence and structure of a new coiled coil protein from a
                    microtubule bundle in Giardia.
   #accession     S33821
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-857 ##label MAR
      ##cross-references EMBL:X64517; NID:g312670; PID:g312671
SUMMARY          #length 857  #molecular-weight 100583  #checksum 6805

  Query Match             2.8%;  Score 95;  DB 2;  Length 857;
  Best Local Similarity  33.3%;  Pred. No. 1.15e+00;
  Matches    15;  Conservative   16;  Mismatches  11;  Indels   3;  Gaps   3;

Db    482 KENSYNFDQLLEQKQQMRSDLNALREKAADYE-RVDRELRLKDKE 525
          || ::::   || |:: || ::||::|:  |: |  | :: |:|:
Qy     56 KETNFSLASGLEAKDECRSAMEALKQKSL-YNCRCKRGMK-KEKN 98


RESULT   12
ENTRY             S54067     #type complete
TITLE             probable membrane protein YPR042c - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   hypothetical protein YP3085.06c; hypothetical protein
                    YP9499.01c
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-Jul-1995 #sequence_revision 26-Jul-1996 #text_change
                    12-Dec-1997
ACCESSIONS        S54067; S61063
REFERENCE         S54059
   #authors       Badcock, K.; Churcher, C.M.
   #submission    submitted to the EMBL Data Library, May 1995
   #accession     S54067
      ##molecule_type DNA
      ##residues      1-508 ##label BAD
      ##cross-references EMBL:Z49219; NID:g805025; PID:g805026; MIPS:YPR042c
      ##experimental_source strain AB972
REFERENCE         S61058
   #authors       Badcock, K.; Churcher, C.M.
   #submission    submitted to the EMBL Data Library, November 1995
   #accession     S61063
      ##molecule_type DNA
      ##residues      465-1075 ##label BAW
      ##cross-references EMBL:Z68111; MIPS:YPR042c
```

```
GENETICS
   #map_position 16R
KEYWORDS          transmembrane protein
FEATURE
   711-727             #domain transmembrane #status predicted #label TMM
SUMMARY           #length 1075  #molecular-weight 119507  #checksum 3888

  Query Match             2.8%;  Score 94;  DB 2;  Length 1075;
  Best Local Similarity  30.2%;  Pred. No. 1.56e+00;
  Matches    26;  Conservative   24;  Mismatches  27;  Indels   9;  Gaps   9;

Db    679 SFIFESVLSHFWTIVQNRYGSRAVRACLEADSIITQCQLLTITSLI-IVLSPYLATDTNG 737
          ::|  | |: | |  |    ||:| :||  ::  ::| ||: ::||  |::|    |::: 
Qy    253 NYICRSRLADFFTNCQ-P-ESRSVSSCLK-ENY-ADC-LLAYSGLIGTVMTPNY-IDSSS 306

Db    738 TLLITWLLDTCTLPNKNLILCDKLVN 763
          :  |  | |: : ::|  | |::|
Qy    307 LSVAPWC-D-CSNSGNDLEECLKFLN 330


RESULT   13
ENTRY             S49886     #type complete
TITLE             probable membrane protein YIL123w - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   hypothetical protein YI8277.06
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change
                    06-Feb-1998
ACCESSIONS        S49886
REFERENCE         S49881
   #authors       Hamlyn, N.; Churcher, C.
   #submission    submitted to the EMBL Data Library, November 1994
   #accession     S49886
      ##molecule_type DNA
      ##residues      1-475 ##label HAM
      ##cross-references GB:Z47047; EMBL:Z46833; NID:g603997; PID:g763223;
                          MIPS:YIL123w
GENETICS
   #gene          SGD:SIM1
      ##cross-references SGD:S0001385; MIPS:YIL123w
   #map_position 9L
KEYWORDS          transmembrane protein
FEATURE
   82-98               #domain transmembrane #status predicted #label TMM
SUMMARY           #length 475  #molecular-weight 48070  #checksum 4186

  Query Match             2.7%;  Score 91;  DB 2;  Length 475;
  Best Local Similarity  31.7%;  Pred. No. 3.92e+00;
  Matches    20;  Conservative   14;  Mismatches  26;  Indels   3;  Gaps   3;

Db    294 CKNGYLYRSNTDADYLCEWGVEAAYVVSKLSKGVAICRTDYPG-TENMVIPTYVEGGSSL 352
          ||  |: ||   ||:: :   |:  | | |    | |   |:|    :: | |::: |||
Qy    250 CKTNYICRSRL-ADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPNYIDS-SSL 307

Db    353 PLT 355
          :::
Qy    308 SVA 310


RESULT   14
ENTRY             JC5621     #type complete
TITLE             epidermal growth factor-like protein, T16 - rat
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              09-Oct-1997 #sequence_revision 07-Nov-1997 #text_change
                    12-Dec-1997
ACCESSIONS        JC5621
REFERENCE         JC5621
   #authors       Ozaki, T.; Kondo, K.; Nakamura, Y.; Ichimiya, S.; Nakagawara,
                    A.; Sakiyama, S.
   #journal       Biochem. Biophys. Res. Commun. (1997) 237:245-250
   #title         Interaction of DA41, a DAN-binding protein, with the
                    epidermal growth factor-like protein, S(1-5).
   #accession     JC5621
      ##molecule_type mRNA
      ##residues      1-493 ##label OZA
      ##cross-references DDBJ:D89730; NID:g2429082; PID:d1023127; PID:g2429083
COMMENT      This protein plays a role in the regulation of cell growth by
               interacting with DAN protein through DA41 protein.
KEYWORDS          glycoprotein
FEATURE
   1-17                #domain signal sequence #status predicted #label SIG\
   28-70,158-199,
   200-237,238-277,
   278-318,319-359     #region epidermal growth factor-like repeat #status
                         predicted\
   249                 #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 493  #molecular-weight 54596  #checksum 6779

  Query Match             2.7%;  Score 92;  DB 2;  Length 493;
  Best Local Similarity  25.6%;  Pred. No. 2.89e+00;
  Matches    10;  Conservative   13;  Mismatches  14;  Indels   2;  Gaps   2;

Db    163 YEQSEHNVCQDIDECTSGTHNCRLDQVCINLRGSFTCHC 201
          ::| || :    ::|  ::  | ||::| : |:::   |
Qy    142 FQQVEH-IPKG-NNCLDAAKACNLDDICKKYRSAYITPC 178


RESULT   15
ENTRY             VCMVM7     #type complete
TITLE             env polyprotein - baboon endogenous virus (strain M7)
ALTERNATE_NAMES   coat polyprotein
CONTAINS          coat protein gp70; coat protein p20E
ORGANISM          #formal_name baboon endogenous virus
   #note          host Papio sp. (baboon)
DATE              30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
                    05-Sep-1997
ACCESSIONS        JT0262
REFERENCE         JT0260
   #authors       Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano,
                    T.
   #journal       Jpn. J. Genet. (1987) 62:127-137
   #title         The entire nucleotide sequence of baboon endogenous virus
                    DNA: a chimeric genome structure of murine type C and
                    simian type D retroviruses.
   #accession     JT0262
      ##molecule_type DNA
      ##residues      1-563 ##label KAT
      ##cross-references GB:M16550; NID:g509586; PID:g332599
GENETICS
   #gene          env
CLASSIFICATION    #superfamily type C retrovirus env polyprotein
KEYWORDS          coat protein; glycoprotein; polyprotein; transmembrane
                    protein
FEATURE
   1-376               #product coat protein gp70 #status predicted #label CGP\
   377-563             #product coat protein p20E #status predicted #label CPP\
   437-449             #region immunosuppressive peptide #status predicted\
   113,219,229,264,
   282,292,306,312,
   321,339,469         #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 563  #molecular-weight 61879  #checksum 3481

  Query Match             2.7%;  Score 92;  DB 1;  Length 563;
  Best Local Similarity  21.4%;  Pred. No. 2.89e+00;
  Matches    24;  Conservative   32;  Mismatches  48;  Indels   8;  Gaps   8;

Db    103 YTSYQQCRSGNKTYYT-ATLLKTQTGGTSDVQVLGSTNKLIQSPCN-GIKGQSICWSTTA 160
          | : :|| :|: | :: |: | :     | :: |   : |    |: |:| :  |
Qy     46 YRTLRQCVAGKETNFSLASGLEAKDECRSAMEAL-KQKSLYNCRCKRGMKKEKNCLRIYW 104

Db    161 PIHVS-DGGGPLDTTRIKSVQRKLEEIHKAL-Y-PELQYHPLA-IPKVRDNL 208
          ::  | :|   |: :   :|  :| :| : : : ::: :: :  |||  : |
Qy    105 SMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDV-FQQVEHIPKGNNCL 155
```

Search completed: Tue Jun 23 18:25:42 1998
Job time : 123 secs.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Tue Jun 23 18:19:39 1998; MasPar time 15.99 Seconds
729.505 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-2
Description: (1-465) from US08866354.pep
Perfect Score: 3369
Sequence: 1 MFLATLYFALPLLDLLLSAE.....PLLVLVVTALSTLLSLTETS 465

Scoring table: PAM 150
Gap 11

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 47.888; Variance 65.919; scale 0.726

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3335	99.0	464	1	GDNR_HUMAN	GDNF RECEPTOR ALPHA PR	0.00e+00
2	3184	94.5	468	1	GDNR_RAT	GDNF RECEPTOR ALPHA PR	0.00e+00
3	3183	94.5	468	1	GDNR_MOUSE	GDNF RECEPTOR ALPHA PR	0.00e+00
4	2593	77.0	469	1	GDNR_CHICK	GDNF RECEPTOR ALPHA PR	0.00e+00
5	1522	45.2	463	1	NRTR_MOUSE	NEURTURIN RECEPTOR ALP	0.00e+00
6	1522	45.2	464	1	NRTR_HUMAN	NEURTURIN RECEPTOR ALP	0.00e+00
7	1480	43.9	465	1	NRTR_CHICK	NEURTURIN RECEPTOR ALP	0.00e+00
8	102	3.0	249	1	ULA3_HCMVA	PROTEIN UL103.	8.87e-03
9	101	3.0	352	1	BRB1_RABIT	B1 BRADYKININ RECEPTOR	1.32e-02
10	93	2.8	169	1	SECB_HAEIN	PROTEIN-EXPORT PROTEIN	2.92e-01
11	96	2.8	504	1	GAG_HV1MA	GAG POLYPROTEIN (CONTA	9.39e-02
12	93	2.8	600	1	SP96_DICDI	SPORE COAT PROTEIN SP9	2.92e-01
13	95	2.8	706	1	YNC1_YEAST	HYPOTHETICAL 80.1 KD P	1.38e-01
14	96	2.8	826	1	MS2_MOUSE	CELL SURFACE ANTIGEN M	9.39e-02
15	95	2.8	857	1	MEDB_GIALA	MEDIAN BODY PROTEIN.	1.38e-01
16	91	2.7	388	1	GAG_HV1W2	GAG POLYPROTEIN (CONTA	6.11e-01
17	91	2.7	475	1	SIM1_YEAST	SIM1 PROTEIN.	6.11e-01
18	91	2.7	503	1	GAG_HV1JR	GAG POLYPROTEIN (CONTA	6.11e-01
19	92	2.7	563	1	ENV_BAEVM	ENV POLYPROTEIN PRECUR	4.23e-01
20	92	2.7	688	1	C1S_HUMAN	COMPLEMENT C1S COMPONE	4.23e-01
21	92	2.7	695	1	CASP_MESAU	CALCIUM-DEPENDENT SERI	4.23e-01
22	86	2.6	167	1	Y491_METJA	HYPOTHETICAL PROTEIN M	3.62e+00
23	86	2.6	316	1	YNF8_YEAST	HYPOTHETICAL 35.0 KD P	3.62e+00
24	89	2.6	327	1	UPAR_MOUSE	UROKINASE PLASMINOGEN	1.26e+00
25	87	2.6	342	1	E13B_BRACM	GLUCAN ENDO-1,3-BETA-G	2.56e+00
26	89	2.6	491	1	CC1_YEAST	CELL DIVISION CONTROL	1.26e+00
27	88	2.6	494	1	Y663_METJA	HYPOTHETICAL PROTEIN M	1.80e+00
28	86	2.6	501	1	GAG_HV1A2	GAG POLYPROTEIN (CONTA	3.62e+00
29	88	2.6	506	1	GAG_HV1MN	GAG POLYPROTEIN (CONTA	1.80e+00
30	86	2.6	511	1	GAG_HV1PV	GAG POLYPROTEIN (CONTA	3.62e+00
31	86	2.6	511	1	GAG_HV1B5	GAG POLYPROTEIN (CONTA	3.62e+00
32	86	2.6	511	1	GAG_HV1B1	GAG POLYPROTEIN (CONTA	3.62e+00
33	87	2.6	514	1	GUXC_FUSOX	PUTATIVE EXOGLUCANASE	2.56e+00
34	88	2.6	521	1	GAG_HV2ST	GAG POLYPROTEIN (CORE	1.80e+00
35	87	2.6	521	1	GAG_HV2BE	GAG POLYPROTEIN (CORE	2.56e+00
36	87	2.6	521	1	GAG_HV2D1	GAG POLYPROTEIN (CORE	2.56e+00
37	88	2.6	522	1	GAG_HV2G1	GAG POLYPROTEIN (CORE	1.80e+00
38	87	2.6	535	1	PPB3_HUMAN	ALKALINE PHOSPHATASE,	2.56e+00
39	87	2.6	563	1	YAS8_SCHPO	PUTATIVE 64.5 KD TRANS	2.56e+00
40	86	2.6	646	1	VE1_HPV07	REPLICATION PROTEIN E1	3.62e+00
41	87	2.6	1182	1	ABL2_HUMAN	TYROSINE-PROTEIN KINAS	2.56e+00
42	88	2.6	1403	1	PRO_DROME	PROTEIN PROSPERO.	1.80e+00
43	87	2.6	1409	1	HAP1_HAEIN	ADHESION AND PENETRATI	2.56e+00
44	87	2.6	2531	1	NTC1_RAT	NEUROGENIC LOCUS NOTCH	2.56e+00
45	86	2.6	2531	1	NTC1_MOUSE	NEUROGENIC LOCUS NOTCH	3.62e+00

ALIGNMENTS

```
RESULT    1
ID   GDNR_HUMAN     STANDARD;      PRT;   464 AA.
AC   P56159;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE   NEUROTROPHIC FACTOR RECEPTOR 1).
GN   GDNFRA OR TRNR1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SUBSTANTIA NIGRA;
RX   MEDLINE; 96270513.
RA   JING S., WEN D., YU Y., HOLST P.L., LUO Y., FANG M., TAMIR R.,
RA   ANTONIO L., HU Z., CUPPLES R., LOUIS J.-C., HU S., ALTROCK B.W.,
RA   FOX G.M.;
RL   CELL 85:1113-1124(1996).
CC   -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC       WITH THE DISULFIDE-LINKED GDNF DIMER AND 2 MOLECULES OF RET (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   MIM; 601496; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GDNF RECEPTOR ALPHA.
FT   PROPEP        ?    464       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   DOMAIN      361    368       POLY-THR.
FT   CARBOHYD     59     59       POTENTIAL.
FT   CARBOHYD    346    346       POTENTIAL.
FT   CARBOHYD    405    405       POTENTIAL.
SQ   SEQUENCE   464 AA;  51291 MW;  2C8C3574 CRC32;

  Query Match           99.0%;  Score 3335;  DB 1;  Length 464;
  Best Local Similarity  99.6%;  Pred. No. 0.00e+00;
  Matches   463;  Conservative    0;  Mismatches   1;  Indels   1;  Gaps   1;

Db      1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy       1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db     181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db     241 PNCL-LQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 299
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db     300 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAPPV 359
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Qy     301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db     360 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 419
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db     420 EKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLTETS 464
           |||||||||||||||||||||||||||||||||||||||||||||
Qy     421 EKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLTETS 465
```

RESULT 2
ID GDNR_RAT STANDARD; PRT; 468 AA.
AC Q62997;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE NEUROTROPHIC FACTOR RECEPTOR 1).
GN GDNFRA OR TRNR1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE; 96270513.
JING S., WEN D., YU Y., HOLST P.L., LUO Y., FANG M., TAMIR R.,
ANTONIO L., HU Z., CUPPLES R., LOUIS J.-C., HU S., ALTROCK B.W.,
FOX G.M.;
RL CELL 85:1113-1124(1996).
CC -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, BRAIN AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR EMBL; U59486; G1399863; -.
KW RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 ? GDNF RECEPTOR ALPHA.
FT PROPEP ? 468 HYDROPHOBIC, REMOVED DURING MATURATION
FT (POTENTIAL).
FT DOMAIN 362 369 POLY-THR.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 347 347 POTENTIAL.
FT CARBOHYD 406 406 POTENTIAL.
SQ SEQUENCE 468 AA; 51649 MW; 6A7A2B2A CRC32;

Query Match 94.5%; Score 3184; DB 1; Length 468;
Best Local Similarity 92.9%; Pred. No. 0.00e+00;
Matches 430; Conservative 27; Mismatches 5; Indels 1; Gaps 1;

```
Db       1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
           ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy       1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
           ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
           |||||||||||| ||||||||||:|||||||||||||||| ||||||||||||||||||||
Qy     121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240
           |:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Qy     181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
           ||||:|||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Qy     241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
           |:|||||||||||||||||:|||||||||||||||||||||||||||||:|||| ||
Qy     301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420
           ||||||||||:|||||||||||||||||||||||||||||||||||:||||:|::::
Qy     361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVS 463
           |:|| ||||||||||||||:||:| ||::|||::|||::
Qy     421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462
```

RESULT 3
ID GDNR_MOUSE STANDARD; PRT; 468 AA.
AC P97785;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE NEUROTROPHIC FACTOR RECEPTOR 1).
GN GDNFRA OR TRNR1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION;
RA WATABE K.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC SIMILARITY).
CC -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR EMBL; AB000800; G1816442; -.
KW RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 ? GDNF RECEPTOR ALPHA.
FT PROPEP ? 468 HYDROPHOBIC, REMOVED DURING MATURATION
FT (POTENTIAL).
FT DOMAIN 362 369 POLY-THR.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 347 347 POTENTIAL.
FT CARBOHYD 406 406 POTENTIAL.
SQ SEQUENCE 468 AA; 51782 MW; 6C64C182 CRC32;

Query Match 94.5%; Score 3183; DB 1; Length 468;

```
  Best Local Similarity  92.9%;  Pred. No. 0.00e+00;
  Matches   430;  Conservative   23;  Mismatches   9;  Indels   1;  Gaps   1;

Db      1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          |||||||| ||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
          ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
          |||||||||||||| ||||||||||||||:|||||||||||||||| ||||||||||||||
Qy    121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 240
          |:||:|||||||||||||||||||||||||||||:|||||||||||||||||||||||||:
Qy    181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
          |||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Qy    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db    301 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
          |||||||||||||||||||||:|||||||||||||||||||||||||||||:|||| ||
Qy    301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db    361 QTTTAMTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 420
          ||||| ||||:|:|||||||||||||||||||||||||||||||||||:||||:|: :|
Qy    361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db    421 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVS 463
            |:|| ||||||||||||||||||:| |:| |||::|||::
Qy    421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462


RESULT    4
ID    GDNR_CHICK     STANDARD;      PRT;   469 AA.
AC    O13156;
DT    01-NOV-1997 (REL. 35, CREATED)
DT    01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT    01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE    GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE    NEUROTROPHIC FACTOR RECEPTOR 1).
GN    GDNFRA.
OS    GALLUS GALLUS (CHICKEN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
      GALLIFORMES.
      [1]
      SEQUENCE FROM N.A.
RA    BUJ-BELLO A., ADU J., PINON L.G., HORTON A., THOMPSON J.,
RA    ROSENTHAL A., CHINCHETRU M., BUCHMAN V.L., DAVIES A.M.;
RL    NATURE 387:721-724(1997).
CC    -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC        AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC        SIMILARITY).
CC    -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC        WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET
CC        (BY SIMILARITY).
CC    -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC        SIMILARITY).
CC    -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR    EMBL; U90541; G2213803; -.
KW    RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT    SIGNAL        1     27       POTENTIAL.
FT    CHAIN        28      ?       GDNF RECEPTOR ALPHA.
FT    PROPEP        ?    469       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                 (POTENTIAL).
FT    DOMAIN      361    368       POLY-THR.
FT    CARBOHYD     62     62       POTENTIAL.
FT    CARBOHYD    163    163       POTENTIAL.
FT    CARBOHYD    346    346       POTENTIAL.
FT    CARBOHYD    405    405       POTENTIAL.

SQ    SEQUENCE   469 AA;  52043 MW;  C8D241C9 CRC32;

  Query Match           77.0%;  Score 2593;  DB 1;  Length 469;
  Best Local Similarity  78.6%;  Pred. No. 0.00e+00;
  Matches   371;  Conservative   44;  Mismatches   43;  Indels   14;  Gaps  10;

Db      1 MFLALLYLALPLADVLLSAEVSGLPGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKE 60
          |||| ||:|||| |:||||||||   ||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLLSAEVSG---GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKE 57

Db     61 SNFSRATGLEAKDECKSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLE 120
          :||| |:||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy     58 TNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLE 117

Db    121 DSPYEPVNSRLSDIFRLAPIVS-V-EPV--LSKGNNCLDAAKACNLNDTCKRFRSAYITP 176
          |||||||||||||||: |::| | : |  ::||||||||||||||:| ||::||||||||
Qy    118 DSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITP 177

Db    177 CTSSTSNEICNKRKCHKALRLFFDKVPPKHSYGMLFCSCRDVACTERRRQTIVPVCSYED 236
          ||:| ||::||:|||||||| ||||||:||||||||||||:|||||||||||||||||:
Qy    178 CTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEE 237

Db    237 REKPNCLNLQESCKKNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVM 296
          ||||||||||:||| ||||||||||||||||||||||||||||||||||||||||||||||
Qy    238 REKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVM 297

Db    297 TPNYIDSSSLSVAPWCDCSNSGNDIDECRKFLNFFQDNTCLKNAIQAFGNGTDVNVWQPI 356
          |||||||||||||||||||||::|| |||||| ||||||||||||||:|| ||||
Qy    298 TPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPA 357

Db    357 LPVQTTTATTTTASRLKNT--GSETTNNEIPTHNDSPACANLQAQKKRKSNESVDTELCL 414
          :|||||||||||| |:||   |:  ::|||||   |:||||||| ||| | :| ||:
Qy    358 FPVQTTTATTTTALRVKNKPLGPAGSENEIPTHV-LPPCANLQAQKL-KSNVSGNTHLCI 415

Db    415 NENAIGKDNTPGVSTSHISSENSFALPTSFYPSTPLILMTIALSLFLFLSSS 466
          ::    |:   | | |||:: :| | | |    |  |:|:  ||| :| |: :
Qy    416 SNGNYEKEGL-GAS-SHITT-KSMAAPPSCGLSPLLVLVVTALSTLLSLTET 464


RESULT    5
ID    NRTR_MOUSE     STANDARD;      PRT;   463 AA.
AC    O08842;
DT    01-NOV-1997 (REL. 35, CREATED)
DT    01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT    01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE    NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (TGF-
DE    BETA RELATED NEUROTROPHIC FACTOR RECEPTOR 2) (GDNF RECEPTOR BETA)
DE    (GDNFR-BETA).
GN    GDNFRB OR TRNR2.
OS    MUS MUSCULUS (MOUSE).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC    EUTHERIA; RODENTIA.
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 97325791.
RA    BALOH R.H., TANSEY M.G., GOLDEN J.P., CREEDON D.J.,
RA    HEUCKEROTH R.O., KECK C.L., ZIMONJIC D.B., POPESCU N.C.,
RA    JOHNSON E.M., MILBRANDT J.;
RL    NEURON 18:793-802(1997).
CC    -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC        AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC    -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC        SIMILARITY).
CC    -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR    EMBL; AF002701; G2145082; -.
KW    RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT    SIGNAL        1      ?       POTENTIAL.
FT    CHAIN         ?      ?       TGF-BETA RELATED NEUROTROPHIC FACTOR
FT                                 RECEPTOR 2.
FT    PROPEP        ?    463       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                 (POTENTIAL).
FT    CARBOHYD     52     52       POTENTIAL.
FT    CARBOHYD    357    357       POTENTIAL.
```

```
FT   CARBOHYD    413    413       POTENTIAL.
SQ   SEQUENCE   463 AA;  51598 MW;  0A2165C0 CRC32;

  Query Match           45.2%;  Score 1522;  DB 1;  Length 463;
  Best Local Similarity  51.3%;  Pred. No. 0.00e+00;
  Matches   203;  Conservative   91;  Mismatches  90;  Indels  12;  Gaps  10;

Db      37 QVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN-TM---LANK-ECQAALEVLQESPLY 91
           ::|||:|:: |  | :||::|||||||:||:: | ::   |  | ||:::|:| | : :||
Qy      26 RLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLY 85

Db      92 DCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEASPYEPVTSRLSDIFRLASIFSGTGAD 151
           :|||||||||||  ||:|||||:  :| :|::: | |||||| |||||||| :::|    :
Qy      86 NCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQ 144

Db     152 PVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICNREISPTERCNRRKCHKALRQFFDRV 211
           :|:|:|||||||||||:| ||| ||:||: |   :|  : |||||||||||||||:|
Qy     145 VEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTTSVS-NDVCNRRKCHKALRQFFDKV 203

Db     212 PSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDKEKPNCLDLRSLCRTDHLCRSRLADF 271
           |:  :| ||||||:| ||:||||||:| ||||::||||||:|:  |:|: :||||||||
Qy     204 PAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADF 263

Db     272 HANCRASYRTITSCPADNYQACLGSYAGMIGFDMTPNYVDSNPTGIVVSPWCNCRGSGNM 331
           :||::  |:::||  :||  || :|:|:||  |||||:||: : : |:|||:|  |||
Qy     264 FTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPNYIDSS-S-LSVAPWCDCSNSGND 321

Db     332 EEECEKFLKDFTENPCLRNAIQAFGNGTDVNM-SPKGPTFSATQAPRVE-KTPSLPDDLS 389
           |||  |||: | :| ||:|||||||||:|| :  |  |  ::| :     :  : |   :
Qy     322 LEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPA 381

Db     390 DSTS-LGTSVITTCTSIQEQGLKANNSKELSMCFTE 424
           | : : | |:  |:::| | ||:| | :  :|:::
Qy     382 GSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISN 417


RESULT    6
ID   NRTR_HUMAN     STANDARD;      PRT;   464 AA.
AC   O00451;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (TGF-
DE   BETA RELATED NEUROTROPHIC FACTOR RECEPTOR 2) (GDNF RECEPTOR BETA)
DE   (GDNFR-BETA).
GN   GDNFRB OR TRNR2.
OS   HOMO SAPIENS (HUMAN).
     EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
     EUTHERIA; PRIMATES.
     [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97325791.
RA   BALOH R.H., TANSEY M.G., GOLDEN J.P., CREEDON D.J.,
RA   HEUCKEROTH R.O., KECK C.L., ZIMONJIC D.B., POPESCU N.C.,
RA   JOHNSON E.M., MILBRANDT J.;
RL   NEURON 18:793-802(1997).
CC   -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; AF002700; G2145080; -.
DR   MIM; 601956; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?      ?       TGF-BETA RELATED NEUROTROPHIC FACTOR
FT                                RECEPTOR 2.
FT   PROPEP        ?    464       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   CARBOHYD     52     52       POTENTIAL.
FT   CARBOHYD    357    357       POTENTIAL.
FT   CARBOHYD    413    413       POTENTIAL.
SQ   SEQUENCE   464 AA;  51558 MW;  3C74BBFB CRC32;

  Query Match           45.2%;  Score 1522;  DB 1;  Length 464;
  Best Local Similarity  48.3%;  Pred. No. 0.00e+00;
  Matches   223;  Conservative  100;  Mismatches 122;  Indels  17;  Gaps  15;

Db      12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
           ||     :|:    | ::|: |    :|||:|:: |  | :||::|||||||:||:: |
Qy       2 FLATLYFALPLLDLLLSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEAS 126
           ::   |  | ||:::|:| | : :||:|||||||||  ||:||||:  :| :|::: | |
Qy      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db     127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
           ||||| ||||||||: :::|    :    :|:|:|||||||||:| ||| ||:||: |
Qy     120 PYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCT 179

Db     187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
           :|  : ||||||||||||||||:||:  :| ||||||:| ||:|||||||:| ||||::
Qy     180 TSVS-NDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db     247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
           ||||||:|:  |:|: :||||||| :||::  ::|:||  :||  || :|:|:||  ||
Qy     239 EKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMT 298

Db     307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
           |||:||| : : |:||| |  |||  ||| ||| |  | :| ||:|||||||||:|| |
Qy     299 PNYIDSS-S-LSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQP 356

Db     367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
           : : |:| |      :  : |   : | : : | |:  |:::| | ||:| | :  :|::
Qy     357 AFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS 416

Db     424 ELTTNIIP-GSNKVIKPNSGPSRARPSAALTVLSVLMLKQAL 464
           :   :    |::  |  :|  : | || :|: | ||::  ||
Qy     417 NGNYEKEGLGASSHITTKS-MA-APPSCGLSPLLVLVVT-AL 455


RESULT    7
ID   NRTR_CHICK     STANDARD;      PRT;   465 AA.
AC   O13157;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (GDNF
DE   RECEPTOR BETA) (GDNFR-BETA).
GN   GDNFRB.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BUJ-BELLO A., ADU J., PINON L.G., HORTON A., THOMPSON J.,
RA   ROSENTHAL A., CHINCHETRU M., BUCHMAN V.L., DAVIES A.M.;
RL   NATURE 387:721-724(1997).
CC   -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; U90542; G2213805; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?      ?       NEURTURIN RECEPTOR ALPHA.
FT   PROPEP        ?    465       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   CARBOHYD    355    355       POTENTIAL.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CARBOHYD    412    412       POTENTIAL.
SQ   SEQUENCE   465 AA;  51908 MW;  5CA073E4 CRC32;

  Query Match           43.9%;  Score 1480;  DB 1;  Length 465;
  Best Local Similarity  46.5%;  Pred. No. 0.00e+00;
```

```
  Matches   203;  Conservative   96;  Mismatches 124;  Indels  14;  Gaps  11;

Db      38 VDCIRANKLCAAEGSCSSRYRTLRQCLAGRDRN-TM---LANK-ECQAALEVLQESPLYD 92
           :|||::|:  |  | |||::|||||||:||:: | ::   |  | ||:::|:| | : :|||:
Qy      27 LDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYN 86

Db      93 CRCKRGMRKEIQCLQVYWSIHLGLAEGEEFYEASPYEPITSRLSDIFRLASIFSGM-DPA 151
           |||||||:||  ||::|||:  :| :|::: | |||||: |||||||| :::| : :
Qy      87 CRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQV 145

Db     152 TN-SKSNHCLDAAKACNLNDNCKRLRSGYISTCSKEISATEHCSRRKCHKALRQFFDNVP 210
           : :|:|:|||||||||||:| ||: ||:||: |:  :|  : |:|||||||||||||:||
Qy     146 EHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTTSVSN-DVCNRRKCHKALRQFFDKVP 204

Db     211 SEYTYRLLFCSCKDQACAEPRRQTIVPFCSYEDKEKPNCLDLRNVCRADHLCRSRLADFH 270
           :  :| :|||||:| ||:| ||||||| ||||::||||||:|:: |::: :||||||||
Qy     205 AKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFF 264

Db     271 ANCQASFQSLTSCPGDNYQACLGSYTGLIGFDMTPNYVDASTTSITISPWCSCKGSGNLE 330
           :|||:  :|:::||  :||  || :|:|||| |||||:|:|  |::::||| |  |||
Qy     265 TNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPNYIDSS--SLSVAPWCDCSNSGNDL 322

Db     331 EECEKFLRDFTENPCLRNAIQAFGNGTDVNL-SPKNPSPPITMLPKVE-KSPALPDDIND 388
           ||| |||  | :| ||:|||||||||:|| :  |  |    |       :    |
Qy     323 EECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAG 382

Db     389 SNTMYDTSIITTCTSIQEHGQKLNKSKEQSLCYSETQLTTDTMPDQKTFVDQKAAGSRHR 448
           |:    | ::  |:::| :  | | | :  || |:     : :     :     |::
Qy     383 SENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSC 442

Db     449 A-ARILPAVPIVLLKLL 464
           : : :|  |   |  ||
Qy     443 GLSPLLVLVVTALSTLL 459


RESULT    8
ID   ULA3_HCMVA     STANDARD;      PRT;   249 AA.
AC   P16734;
DT   01-AUG-1990 (REL. 15, CREATED)
DT   01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT   01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE   PROTEIN UL103.
GN   UL103.
OS   HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
     MEDLINE; 90269039.
     CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
     HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA   PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RL   CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC       EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.
DR   EMBL; X17403; G59711; -.
DR   PIR; S09868; S09868.
SQ   SEQUENCE   249 AA;  28636 MW;  239ADB2D CRC32;

  Query Match             3.0%;  Score 102;  DB 1;  Length 249;
  Best Local Similarity  36.2%;  Pred. No. 8.87e-03;
  Matches    17;  Conservative    8;  Mismatches  20;  Indels   2;  Gaps   2;

Db     106 CTVISCVENCNLTRKCLHDLLQYLDAVNVRESFGRLLHHSARRLICS 152
           ||     : ||  |||   | |::| |  : |:| :|  | | : |:
Qy     178 CTTSVSNDVCN-RRKCHKALRQFFDKVPAKHSYG-MLFCSCRDIACT 222


RESULT    9
ID   BRB1_RABIT     STANDARD;      PRT;   352 AA.
AC   P48748;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   B1 BRADYKININ RECEPTOR (BK-1 RECEPTOR).
GN   BDKRB1.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; LAGOMORPHA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=AORTA;
RX   MEDLINE; 96085127.
RA   MACNEIL T., BIERILO K.K., MENKE J.G., HESS J.F.;
RL   BIOCHIM. BIOPHYS. ACTA 1264:223-228(1995).
CC   -!- FUNCTION: THIS IS A RECEPTOR FOR BRADYKININ. COULD BE A FACTOR IN
CC       CHRONIC PAIN AND INFLAMMATION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- INDUCTION: THE B1 RECEPTOR IS SYNTHESIZED DE NOVO IN RESPONSE TO
CC       TISSUE INJURY.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; U20507; G1041821; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     39       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     40     63       1 (POTENTIAL).
FT   DOMAIN       64     72       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     73     97       2 (POTENTIAL).
FT   DOMAIN       98    110       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    111    132       3 (POTENTIAL).
FT   DOMAIN      133    154       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    177       4 (POTENTIAL).
FT   DOMAIN      178    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    225       5 (POTENTIAL).
FT   DOMAIN      226    246       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    271       6 (POTENTIAL).
FT   DOMAIN      272    290       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    291    314       7 (POTENTIAL).
FT   DOMAIN      315    352       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    109    188       BY SIMILARITY.
FT   LIPID       329    329       PALMITATE (POTENTIAL).
FT   CARBOHYD     13     13       POTENTIAL.
FT   CARBOHYD     21     21       POTENTIAL.
FT   CARBOHYD    184    184       POTENTIAL.
SQ   SEQUENCE   352 AA;  39503 MW;  70F06F62 CRC32;

  Query Match             3.0%;  Score 101;  DB 1;  Length 352;
  Best Local Similarity  26.0%;  Pred. No. 1.32e-02;
  Matches    26;  Conservative   32;  Mismatches  32;  Indels  10;  Gaps  10;

Db     216 AILFFNCHILA-SLRRRGERVPSRCG-GPRDSKSTALIL-TLVASFLVCWAPYHFFAFLE 272
           ::|| :|: :| : |||   ||  |:   |: |:  | |     :  :|::    | | :
Qy     210 GMLFCSCRDIACTERRRQTIVPV-CSYEERE-KPNCLNLQDSCKTNYICRSRLADF-FTN 266

Db     273 CLWQVHAIGGCFWEEFTDLGLQLSNFSAFVNSCLNPVIYV 312
           |  : ::::::|: |:::|  | |: :|::: : : |  |:
Qy     267 CQPESRSVSSCLKENYAD-CL-LA-YSGLIGTVMTPN-YI 302


RESULT   10
ID   SECB_HAEIN     STANDARD;      PRT;   169 AA.
AC   P44853;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROTEIN-EXPORT PROTEIN SECB.
GN   SECB OR HI0743.
OS   HAEMOPHILUS INFLUENZAE.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   PASTEURELLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
```

```
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RL   SCIENCE 269:496-512(1995).
CC   -!- FUNCTION: THIS IS ONE OF THE PROTEINS REQUIRED FOR THE NORMAL
CC       EXPORT OF ENVELOPE PROTEINS OUT OF THE CELL CYTOPLASM; IT MAY
CC       BE INVOLVED IN THE INITIATION OF THE EXPORTING PROCESS, BY BINDING
CC       TO THE NASCENT POLYPEPTIDE VIA A SIGNAL SEQUENCE, MAINTAINING
CC       A STABLE AND PRE-TRANSLOCATION CONFORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC       COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       (BY SIMILARITY).
DR   EMBL; U32758; G1573750; -.
DR   TIGR; HI0743; -.
KW   PROTEIN TRANSPORT; TRANSLOCATION.
SQ   SEQUENCE   169 AA;  19132 MW;  74BB8A0E CRC32;

  Query Match             2.8%;  Score 93;  DB 1;  Length 169;
  Best Local Similarity  25.0%;  Pred. No. 2.92e-01;
  Matches    16;  Conservative   17;  Mismatches  30;  Indels   1;  Gaps   1;

Db      75 LEDSGDVAFICEVKQAGVFTISGLEDVQMAHCLTSQCPNMLFPYARELVSNLVNRGTFPA 134
           |:||    :||  : |  ||    |   :: ||    :: |::|:  |:: ::    : :
Qy     246 LQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYS-GLIGTVMTPNYIDS 304

Db     135 LNLS 138
           :||
Qy     305 SSLS 308


RESULT   11
ID   GAG_HV1MA      STANDARD;      PRT;   504 AA.
AC   P04594;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   GAG POLYPROTEIN (CONTAINS: CORE PROTEINS P17, P24, P2, P7, P1, P6).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (MAL ISOLATE) (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
     MEDLINE; 86245056.
     ALIZON M., WAIN-HOBSON S., MONTAGNIER L., SONIGO P.;
     CELL 46:63-74(1986).
CC   -!- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC       ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC       REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC       MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC       RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC       GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC       BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC   -!- PTM: THE P24 PROTEIN IS PHOSPHORYLATED.
DR   EMBL; K03456; G328021; -.
DR   EMBL; X04415; G60229; -.
DR   EMBL; A07116; G492872; -.
DR   HSSP; P03348; 1HVN.
DR   HIV; K03456; GAG$MAL.
KW   AIDS; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION; PHOSPHORYLATION.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   CHAIN         1    137       CORE PROTEIN P17 (MATRIX PROTEIN).
FT   CHAIN       138    368       CORE PROTEIN P24 (CORE ANTIGEN).
FT   CHAIN       369    383       CORE PROTEIN P2.
FT   CHAIN       384    437       CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT   CHAIN       438    453       CORE PROTEIN P1.
FT   CHAIN       454    504       CORE PROTEIN P6.
SQ   SEQUENCE   504 AA;  56001 MW;  4500CD44 CRC32;

  Query Match             2.8%;  Score 96;  DB 1;  Length 504;
  Best Local Similarity  41.9%;  Pred. No. 9.39e-02;
  Matches    18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db      27 KYR-LKHLVWASRELERFALNPGLLETGEGCQQIMEQL-QSTL 67
           ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy      45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   12
ID   SP96_DICDI     STANDARD;      PRT;   600 AA.
AC   P14328;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   SPORE COAT PROTEIN SP96.
GN   COTA.
OS   DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC   EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; SARCODINA; RHIZOPODA;
OC   EUMYCETOZOA; DICTYOSTELIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE; 90067962.
RA   FOSNAUGH K., LOOMIS W.F.;
RL   NUCLEIC ACIDS RES. 17:9489-9489(1989).
CC   -!- SUBCELLULAR LOCATION: OUTER LAYER OF THE COAT MATRIX AND
CC       INTERSPORE MATRIX.
CC   -!- PTM: PHOSPHORYLATED AND FUCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED,
CC       MAY CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES.
CC   -!- SIMILARITY: CONTAINS 4 PRESPORE MOTIFS.
DR   EMBL; X16491; G295736; -.
DR   PIR; S07638; S07638.
DR   DICTYDB; DD03007; COTA.
KW   GLYCOPROTEIN; PHOSPHORYLATION; REPEAT; SPORULATION.
FT   REPEAT      185    197       PRESPORE MOTIF 1.
FT   REPEAT      221    233       PRESPORE MOTIF 2.
FT   REPEAT      298    310       PRESPORE MOTIF 3.
FT   REPEAT      395    407       PRESPORE MOTIF 4.
SQ   SEQUENCE   600 AA;  59589 MW;  B18FCDB6 CRC32;

  Query Match             2.8%;  Score 93;  DB 1;  Length 600;
  Best Local Similarity  40.4%;  Pred. No. 2.92e-01;
  Matches    19;  Conservative    8;  Mismatches  14;  Indels   6;  Gaps   6;

Db     298 CRNIQCPTGYRCEDHNRNPICVLEERENPDNCLTCNDVNCEASGLVC 344
           ||:| | |  |   :   |:|  ||||:| |||   | :| ::  :|
Qy     216 CRDIAC-TE-R-RRQTIVPVCSYEEREKP-NCLNLQD-SCKTN-YIC 256


RESULT   13
ID   YNC1_YEAST     STANDARD;      PRT;   706 AA.
AC   P53973;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 80.1 KD PROTEIN IN UME3-PUB1 INTERGENIC REGION.
GN   YNL021W OR N2819.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ANDRE B., IRAQUI HOUSSAINI I., URRESTARAZU L.A., VISSERS S.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: BELONGS TO THE RPD3/C08B11.2/ACUC FAMILY.
DR   EMBL; Z71297; E239664; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   706 AA;  80069 MW;  04B8BCAD CRC32;

  Query Match             2.8%;  Score 95;  DB 1;  Length 706;
  Best Local Similarity  26.6%;  Pred. No. 1.38e-01;
  Matches    21;  Conservative   23;  Mismatches  31;  Indels   4;  Gaps   3;
```

```
Db    330 LVIISSGFDAADGDTIGQCHVTPSCYGHMTHMLKSLARGN--LCVVLEGGYNLDAIARSA 387
          | : :  : :| ::     || |:|    :::|||   ||  ||:  :| |: :::: |:
Qy    371 LRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS-NGNYEKEGLGASS 429

Db    388 LSVAKVLIGEPPDELPDPL 406
            :| : : |:  |: ||
Qy    430 HITTKSMAAPPSCGLS-PL 447


RESULT   14
ID   MS2_MOUSE      STANDARD;      PRT;   826 AA.
AC   Q05910;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CELL SURFACE ANTIGEN MS2 PRECURSOR (EC 3.4.24.-) (MACROPHAGE CYSTEINE-
DE   RICH GLYCOPROTEIN).
GN   ADAM8 OR MS2.
     MUS MUSCULUS (MOUSE).
     EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
     EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RA   YAMAMOTO S., YOSHIYAMA K., SETOGUCHI M., MATSUURA K., HIGUCHI Y.,
RA   AKIZUKI S.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE; 91197896.
RA   YOSHIDA S., SETOGUCHI M., HIGUCHI Y., AKIZUKI S., YAMAMOTO S.;
RL   INT. IMMUNOL. 2:585-591(1990).
CC   -!- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUKOCYTES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MACROPHAGES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
DR   EMBL; X13335; E264744; -.
DR   EMBL; X13335; E225615; -.
DR   MGD; MGI:107825; ADAM8.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW   TRANSMEMBRANE; GLYCOPROTEIN; ANTIGEN; ZINC; HYDROLASE;
KW   METALLOPROTEASE; SIGNAL.
FT   SIGNAL        1    16       POTENTIAL.
FT   CHAIN        17   826       CELL SURFACE ANTIGEN MS2.
FT   DOMAIN       17   658       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    659   683       POTENTIAL.
FT   DOMAIN      684   826       CYTOPLASMIC (POTENTIAL).
FT   METAL       329   329       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    330   330       BY SIMILARITY.
FT   METAL       333   333       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       339   339       ZINC (CATALYTIC) (PROBABLE).
FT   CARBOHYD     89    89       POTENTIAL.
FT   CARBOHYD    260   260       POTENTIAL.
FT   CARBOHYD    431   431       POTENTIAL.
FT   CARBOHYD    614   614       POTENTIAL.
SQ   SEQUENCE   826 AA;  90046 MW;  7C26F36F CRC32;

  Query Match             2.8%;  Score 96;  DB 1;  Length 826;
  Best Local Similarity  33.3%;  Pred. No. 9.39e-02;
  Matches    17;  Conservative   10;  Mismatches  18;  Indels   6;  Gaps   6;

Db    585 ELVLQGTKC-EEGKVC-MDGSCQDLR-VYRSENCSAKCNNHGVCNHKRECH 632
          | :  | :| : :| | :|  |   |  | :  |::  :|  ||| :| ||
Qy    146 EHIPKGNNCLDAAKACNLDDICKKYRSAYITP-CTTSVSND-VCN-RRKCH 193


RESULT   15
ID   MEDB_GIALA     STANDARD;      PRT;   857 AA.
AC   Q08014;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   MEDIAN BODY PROTEIN.
OS   GIARDIA LAMBLIA (GIARDIA INTESTINALIS).
OC   EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; DIPLOMONADIDA;
OC   HEXAMITIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PORTLAND-1;
RX   MEDLINE; 93287123.
RA   MARSHALL J., HOLBERTON D.V.;
RL   J. MOL. BIOL. 231:521-530(1993).
CC   -!- FUNCTION: MAY HAVE A ROLE IN IMMOBILISING THE MICROTUBULES
CC       BETWEEN CELL DIVISIONS.
CC   -!- SUBCELLULAR LOCATION: MEDIAN BODY.
CC   -!- DOMAIN: SHOWS AN ALPHA-HELICAL COILED COIL STRUCTURE (30
CC       REPEATING HEPTADS).
DR   EMBL; X64517; G312671; -.
KW   CYTOSKELETON; MICROTUBULES; COILED COIL; HEPTAD REPEAT PATTERN.
SQ   SEQUENCE   857 AA;  100583 MW;  5E2BBA84 CRC32;

  Query Match             2.8%;  Score 95;  DB 1;  Length 857;
  Best Local Similarity  33.3%;  Pred. No. 1.38e-01;
  Matches    15;  Conservative   16;  Mismatches  11;  Indels   3;  Gaps   3;

Db    482 KENSYNFDQLLEQKQQMRSDLNALREKAADYE-RVDRELRLKDKE 525
          || ::::   || |:: || ::||::|:  |: |  | :: |:|:
Qy     56 KETNFSLASGLEAKDECRSAMEALKQKSL-YNCRCKRGMK-KEKN 98


Search completed: Tue Jun 23 18:20:47 1998
Job time : 68 secs.
```

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 23 18:21:04 1998; MasPar time 28.26 Seconds
692.865 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-2
Description: (1-465) from US08866354.pep
Perfect Score: 3369
Sequence: 1 MFLATLYFALPLLDLLLSAE.....PLLVLVVTALSTLLSLTETS 465

Scoring table: PAM 150
Gap 11

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl5
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_vertebrate
13:sp_unclassified

Statistics: Mean 46.387; Variance 78.836; scale 0.588

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3276	97.2	460	2	O15507	RET LIGAND 1.	0.00e+00
2	3189	94.7	468	10	O35246	GDNF RECEPTOR ALPHA.	0.00e+00
3	3096	91.9	463	10	O35252	GDNF RECEPTOR BETA.	0.00e+00
4	3091	91.7	463	10	O35748	GDNFR-ALPHA/TRNR1-DELT	0.00e+00
5	1525	45.3	464	10	O35977	GLIAL CELL LINE-DERIVE	0.00e+00
6	1522	45.2	464	2	O15316	GLIAL CELL LINE-DERIVE	0.00e+00
7	1522	45.2	464	2	O15328	RET LIGAND 2.	0.00e+00
8	714	21.2	397	10	O35118	GFRALPHA-3.	4.75e-136
9	714	21.2	397	10	O35325	GLIAL CELL LINE-DERIVE	4.75e-136
10	103	3.1	492	11	Q80624	GAG PROTEIN.	8.13e-02
11	100	3.0	129	11	Q97845	MATRIX PROTEIN P17 (FR	2.23e-01
12	100	3.0	130	11	O36786	MA-P17 (FRAGMENT).	2.23e-01
13	100	3.0	130	11	O36808	MA-P17 (FRAGMENT).	2.23e-01
14	100	3.0	133	11	Q97725	MATRIX PROTEIN P17 (FR	2.23e-01
15	100	3.0	486	11	Q77804	GAG PROTEIN.	2.23e-01
16	98	2.9	130	11	O36772	MA-P17 (FRAGMENT).	4.32e-01
17	99	2.9	243	11	P89796	GAG PROTEIN (FRAGMENT)	3.11e-01
18	98	2.9	1127	3	Q94248	CODED FOR BY C. ELEGAN	4.32e-01
19	96	2.8	70	11	P89780	GAG PROTEIN (FRAGMENT)	8.28e-01
20	96	2.8	130	11	Q75890	GAG PROTEIN (FRAGMENT)	8.28e-01
21	96	2.8	130	11	O36773	MA-P17 (FRAGMENT).	8.28e-01
22	96	2.8	130	11	Q75885	GAG PROTEIN (FRAGMENT)	8.28e-01
23	96	2.8	130	11	Q75873	GAG PROTEIN (FRAGMENT)	8.28e-01
24	96	2.8	130	11	Q75884	GAG PROTEIN (FRAGMENT)	8.28e-01
25	96	2.8	130	11	Q75888	GAG PROTEIN (FRAGMENT)	8.28e-01
26	96	2.8	130	11	Q75892	GAG PROTEIN (FRAGMENT)	8.28e-01
27	96	2.8	130	11	Q75887	GAG PROTEIN (FRAGMENT)	8.28e-01
28	96	2.8	130	11	Q76944	GAG PROTEIN (FRAGMENT)	8.28e-01
29	96	2.8	130	11	Q76946	GAG PROTEIN (FRAGMENT)	8.28e-01
30	96	2.8	130	11	Q75866	GAG PROTEIN (FRAGMENT)	8.28e-01
31	96	2.8	130	11	Q75891	GAG PROTEIN (FRAGMENT)	8.28e-01
32	96	2.8	130	11	Q78625	GAG PROTEIN (FRAGMENT)	8.28e-01
33	96	2.8	130	11	Q75864	GAG PROTEIN (FRAGMENT)	8.28e-01
34	96	2.8	130	11	Q75867	GAG PROTEIN (FRAGMENT)	8.28e-01
35	96	2.8	130	11	Q75886	GAG PROTEIN (FRAGMENT)	8.28e-01
36	96	2.8	130	11	Q75868	GAG PROTEIN (FRAGMENT)	8.28e-01
37	96	2.8	130	11	Q75869	GAG PROTEIN (FRAGMENT)	8.28e-01
38	96	2.8	130	11	Q75872	GAG PROTEIN (FRAGMENT)	8.28e-01
39	96	2.8	130	11	Q75870	GAG PROTEIN (FRAGMENT)	8.28e-01
40	96	2.8	130	11	Q75865	GAG PROTEIN (FRAGMENT)	8.28e-01
41	96	2.8	130	11	Q75871	GAG PROTEIN (FRAGMENT)	8.28e-01
42	96	2.8	130	11	Q75883	GAG PROTEIN (FRAGMENT)	8.28e-01
43	96	2.8	133	11	Q97724	MATRIX PROTEIN P17 (FR	8.28e-01
44	96	2.8	304	11	Q77701	GAG PROTEIN.	8.28e-01
45	96	2.8	492	11	Q79351	GAG PROTEIN.	8.28e-01

ALIGNMENTS

RESULT 1
ID O15507 PRELIMINARY; PRT; 460 AA.
AC O15507;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE RET LIGAND 1.
GN RETL1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA PEPINSKY R.B., CATE R.L.;
RL PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA PEPINSKY R.B., CATE R.L.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=SUBSTANTIA NIGRA;
RA HISHIKI T., KONDOH K., ICHIMIYA S., NIMURA Y., SEKI N., OZAKI T.,
RA SAKIYAMA S., TAKAHASHI H., OHNUMA N., TANABE M., FUJIMURA S.,
RA NAKAGAWARA A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U97144; G2282026; -.
DR EMBL; U95847; G2459742; -.
SQ SEQUENCE 460 AA; 50838 MW; 022FFECA CRC32;

Query Match 97.2%; Score 3276; DB 2; Length 460;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 460; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Db 1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

```
Db     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRVVPFIS-----VEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 175
Qy    121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db    176 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 235
Qy    181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db    236 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 295
Qy    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db    296 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 355
      301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

      356 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 415
Qy    361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db    416 EKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLTETS 460
Qy    421 EKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLTETS 465
```

```
RESULT    2
ID   O35246         PRELIMINARY;   PRT;   468 AA.
AC   O35246;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA.
GN   GDNFR-ALPHA.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57; TISSUE=LIVER;
RA   DEY B.K., WONG Y.W., TOO H.P.;
RL   NEUROREPORT 9:0-0(0001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57; TISSUE=LIVER;
RA   DEY B.K., WONG Y.W., TOO H.P.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF014117; G2624961; -.
SQ   SEQUENCE   468 AA;  51752 MW;  AFDCE6A1 CRC32;

  Query Match           94.7%;  Score 3189;  DB 10;  Length 468;
  Best Local Similarity  93.1%;  Pred. No. 0.00e+00;
  Matches   431;  Conservative   23;  Mismatches   8;  Indels   1;  Gaps   1;

Db      1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
Qy      1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
Qy    121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 240
Qy    181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
Qy    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db    301 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
Qy    301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db    361 QTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 420
Qy    361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db    421 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVS 463
Qy    421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462
```

```
RESULT    3
ID   O35252         PRELIMINARY;   PRT;   463 AA.
AC   O35252;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR BETA.
GN   GDNFR-BETA.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57;
RA   DEY B.K., WONG Y.W., TOO H.P.;
RL   NEUROREPORT 9:0-0(0001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57;
RA   DEY B.K., WONG Y.W., TOO H.P.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF015172; G2624963; -.
SQ   SEQUENCE   463 AA;  51134 MW;  910EF17F CRC32;

  Query Match           91.9%;  Score 3096;  DB 10;  Length 463;
  Best Local Similarity  92.0%;  Pred. No. 0.00e+00;
  Matches   426;  Conservative   23;  Mismatches   8;  Indels   6;  Gaps   2;

Db      1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
Qy      1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFIS-----VEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 175
Qy    121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db    176 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 235
Qy    181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db    236 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 295
Qy    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db    296 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 355
Qy    301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db    356 QTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 415
Qy    361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420
```

```
Db    416 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVS 458
          |:|| |||||||||||||||||||:| |:| |||::|||::
Qy    421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462


RESULT    4
ID   O35748         PRELIMINARY;   PRT;   463 AA.
AC   O35748;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GDNFR-ALPHA/TRNR1-DELTA PROTEIN.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HANNOVER;
RA   ZHONG J., ANNIES M., HEUMANN R.;
     SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
     EMBL; AJ002072; E1154274; -.
     SEQUENCE   463 AA;  51032 MW;  93277F91 CRC32;

  Query Match           91.7%;  Score 3091;  DB 10;  Length 463;
  Best Local Similarity  91.8%;  Pred. No. 0.00e+00;
  Matches   425;  Conservative   27;  Mismatches   5;  Indels   6;  Gaps   2;

Db      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
          ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFIS-----VEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 175
          |||||||||||||| |||||     ||||:|||||||||||||||| ||||||||||||||
Qy    121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180

Db    176 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 235
          |:||:||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Qy    181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240

Db    236 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 295
          ||||:||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Qy    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 300

Db    296 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 355
          |:|||||||||||||||||||:|||||||||||||||||||||||||:|||| ||
Qy    301 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 360

Db    356 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 415
          |||||||||:|||||||||||||||||||||||||||||||||||||:||||:|::::
Qy    361 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 420

Db    416 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVS 458
          |:|| ||||||||||||||||:||:| ||::|||::|||::
Qy    421 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 462


RESULT    5
ID   O35977         PRELIMINARY;   PRT;   464 AA.
AC   O35977;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR RECEPTOR-BETA.
GN   GDNFR-BETA OR RETL2.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   TRUPP M., RAYNOSCHEK C., IBANEZ C.F.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN/KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN/KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF005226; G2232252; -.
DR   EMBL; U97143; G2282024; -.
SQ   SEQUENCE   464 AA;  51668 MW;  B1A2BD11 CRC32;

  Query Match           45.3%;  Score 1525;  DB 10;  Length 464;
  Best Local Similarity  49.6%;  Pred. No. 0.00e+00;
  Matches   209;  Conservative   93;  Mismatches 106;  Indels  13;  Gaps  11;

Db     12 FLDETLRSLASPSSLQGSELHGWRPQVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
          ||     :|:    | ::|: |   ::|||:|:: |  | :||::|||||||:||:: |
Qy      2 FLATLYFALPLLDLLLSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEEFYEAS 126
          ::   |  | ||:::|:| | : :|||:|||||||||  ||:||||:  :| :|::: | |
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGTDPAVSTKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
          ||||| |||||||: :::|    :     |:|:||||||||||:| ||| ||:||: |
Qy    120 PYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
            :|  : ||||||||||||||:||:  :| |||||:| ||:|||||||:| ||||::
Qy    180 TSVS-NDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRSLCRTDHLCRSRLADFHANCRASYRTITSCPADNYQACLGSYAGMIGFDMT 306
          ||||||:|  |:|: :|||||||| :|||::  |:::||  :||  || :|:|:||  ||
Qy    239 EKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSNPTGIVVSPWCNCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNM-SP 365
          |||:||: : : |:|||:|  |||  ||| |||  | :| ||:||||||||:|| :  |
Qy    299 PNYIDSS-S-LSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQP 356

Db    366 KGPSLPATQAPRVE-KTPSLPDDLSDSTS-LGTSVITTCTSIQEQGLKANNSKELSMCFT 423
          |   :| :     :  : |   : | : : | |:  |:::| | ||:| | :  :|::
Qy    357 AFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS 416

Db    424 E 424
          :
Qy    417 N 417


RESULT    6
ID   O15316         PRELIMINARY;   PRT;   464 AA.
AC   O15316;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR RECEPTOR BETA.
GN   GDNFR-BETA.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL BRAIN;
RA   WARTIOVAARA K., SUVANTO P., HORELLI-KUITUNEN N., LINDAHL M.,
RA   MOSHNYAKOV M., AIRAKSINEN M.S., PALOTIE A., SARIOLA H., SAARMA M.;
```

```
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U93703; G2228737; -.
SQ   SEQUENCE   464 AA;  51530 MW;  DE80D543 CRC32;

  Query Match           45.2%;  Score 1522;  DB 2;  Length 464;
  Best Local Similarity  48.3%;  Pred. No. 0.00e+00;
  Matches   223;  Conservative  100;  Mismatches 122;  Indels  17;  Gaps  15;

Db     12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy      2 FLATLYFALPLLDLLLSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEEFYEAS 126
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy    120 PYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy    180 TSVS-NDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
Qy    239 EKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
Qy    299 PNYIDSS-S-LSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQP 356

Db    367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
Qy    357 AFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS 416

Db    424 ELTTNIIP-GSNKVIKPNSGPSRARPSAALTVLSVLMLKQAL 464
Qy    417 NGNYEKEGLGASSHITTKS-MA-APPSCGLSPLLVLVVT-AL 455


RESULT    7
ID   O15328         PRELIMINARY;   PRT;   464 AA.
AC   O15328;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   RET LIGAND 2.
GN   RETL2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U97145; G2282028; -.
SQ   SEQUENCE   464 AA;  51543 MW;  288A8BD8 CRC32;

  Query Match           45.2%;  Score 1522;  DB 2;  Length 464;
  Best Local Similarity  48.3%;  Pred. No. 0.00e+00;
  Matches   223;  Conservative  100;  Mismatches 122;  Indels  17;  Gaps  15;

Db     12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy      2 FLATLYFALPLLDLLLSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEEFYEAS 126
Qy     61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy    120 PYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy    180 TSVS-NDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
Qy    239 EKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
Qy    299 PNYIDSS-S-LSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQP 356

Db    367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
Qy    357 AFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCIS 416

Db    424 ELTTNIIP-GSNKVIKPNSGPSRARPSAALTVLSVLMLKLAL 464
Qy    417 NGNYEKEGLGASSHITTKS-MA-APPSCGLSPLLVLVVT-AL 455


RESULT    8
ID   O35118         PRELIMINARY;   PRT;   397 AA.
AC   O35118;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GFRALPHA-3.
GN   GFRALPHA-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL6; TISSUE=HEART;
RA   NOMOTO S., ITO S., YANG L., ADACHI K., KIUCHI K.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB008833; D1024441; -.
SQ   SEQUENCE   397 AA;  44307 MW;  A80E0D24 CRC32;

  Query Match           21.2%;  Score 714;  DB 10;  Length 397;
  Best Local Similarity  35.2%;  Pred. No. 4.75e-136;
  Matches   123;  Conservative   70;  Mismatches 135;  Indels  21;  Gaps  17;

Db     11 LLMILLLVLSLWLPLGAGNSLATENRFVNSCTQARKKCEANPACKAAYQHLGSCTSSLSR 70
Qy      2 FLATLYFALPL-LDLLLSAEVSGGDRL-D-CVKASDQCLKEQSCSTKYRTLRQCVAGKET 58

Db     71 PLPLEES-AMSADCLEAAEQLRNSSLIDCRCHRRMKHQATCLDIYWTVHPARSLGDYELD 129
Qy     59 NFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQ-GNDLLE 117

Db    130 VSPYE--DTVTSKPWKM-N-LSKL-NMLK--PDSDLCLKFAMLCTLHDKCDRLRKAYGEA 182
Qy    118 DSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITP 177

Db    183 CS-GIR---CQRHLCLAQLRSFFEKAAESHAQGLLLCPCAPEDAGCGERRRNTIAPSCAL 238
Qy    178 CTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSC-R-DIACTERRRQTIVPVCSY 235

Db    239 PS-VTPNCLDLRSFCRADPLCRSRLMDFQTHCHPMDI-LGTCATEQ-SRCLRAYLGLIGT 295
```

```
              ||||:|:  |::: :||||| || |:|:|    :::|  |  : || || |||||
Qy    236 EEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGT 295

Db    296 AMTPNFISKVNTTVALSCTCRGSGNLQDECEQLERSFSQNPCLVEAIAA 344
          ||||:|   : :||  | |  |||  :|| :   | :| || :|| |
Qy    296 VMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQA 344


RESULT    9
ID   O35325         PRELIMINARY;   PRT;   397 AA.
AC   O35325;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR FAMILY RECEPTOR ALPHA-3.
GN   GFRA-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   TRUPP M., RAYNOSCHEK C., IBANEZ C.F.;
RL   SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF020305; G2429367; -.
SQ   SEQUENCE   397 AA;  44333 MW;  F0C0C841 CRC32;

  Query Match           21.2%;  Score 714;  DB 10;  Length 397;
  Best Local Similarity  35.2%;  Pred. No. 4.75e-136;
  Matches   123;  Conservative   70;  Mismatches 135;  Indels  21;  Gaps  16;

Db     11 LLMILLLVLSLWLPLGAGNSLATENRFVNSCTQARKKCEANPACKAAYQHLGSCTSSLSR 70
          :|  | : |:| | |  :  ::  :|: : |  |   |  : :| : |: |  | ::
Qy      2 FLATLYFALPL-LDLLLSAEVSGGDRL-D-CVKASDQCLKEQSCSTKYRTLRQCVAGKET 58

Db     71 PLPLEES-AMSADCLEAAEQLRNSSLIDCRCHRRMKHQATCLDIYWTVHPARSLGDYELD 129
          ::|  :    :|  | | |:  || :||| | || :  || |||::  :   |:  |:
Qy     59 NFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQ-GNDLLE 117

Db    130 VSPYE--DTVTSKPWKM-N-LSKL-NMLK--PDSDLCLKFAMLCTLHDKCDRLRKAYGEA 182
          ||||  ::  |   ::   :| :   :   | :: ||  |: | | | | : | ||  :
Qy    118 DSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITP 177

Db    183 CS-GIR---CQRHLCLAQLRSFFEKAAESHAQGLLLCPCPPEDAGCGERRRNTIAPSCAL 238
          |: ::    | |: |   || ||:| :  |: |:|:|:|   | :| |||| || | |:
Qy    178 CTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCR--DIACTERRRQTIVPVCSY 235

Db    239 PS-VTPNCLDLRSFCRADPLCRSRLMDFQTHCHPMDI-LGTCATEQ-SRCLRAYLGLIGT 295
             ||||:|:  |::: :||||| || |:|:|    :::|  |  : || || |||||
      236 EEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGT 295

      296 AMTPNFISKVNTTVALSCTCRGSGNLQDECEQLERSFSQNPCLVEAIAA 344
          ||||:|   : :||  | |  |||  :|| :   | :| || :|| |
Qy    296 VMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQA 344


RESULT   10
ID   Q80624         PRELIMINARY;   PRT;   492 AA.
AC   Q80624;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   GAG PROTEIN.
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VI205; TISSUE=BLOOD;
RA   LOUWAGIE J.J., MCCUTCHAN F., BRENNAN T., PEETERS M., BRENNAN T.,
RA   SANDERS-BUELL E., EDDY G., DER GROEN G., FRANSEN K.,
RA   GERSHY-DAMET M., DELEYS R., BURKE D.;
RL   SUBMITTED (JUN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L11785; G555036; -.
SQ   SEQUENCE   492 AA;  54981 MW;  930293D1 CRC32;

  Query Match           3.1%;  Score 103;  DB 11;  Length 492;
  Best Local Similarity  42.5%;  Pred. No. 8.13e-02;
  Matches   17;  Conservative   9;  Mismatches  10;  Indels   4;  Gaps   4;

Db     26 SKKYR-LKHIVWASRELKRFALNPGLLETSEGCRQIIEQL 64
          | ||| |:: | |:::|   |:| :|| |: : ||  :| |
Qy     43 STKYRTLRQCV-AGKE-TNFSLASGL-EAKDECRSAMEAL 79


RESULT   11
ID   Q97845         PRELIMINARY;   PRT;   129 AA.
AC   Q97845;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE   MATRIX PROTEIN P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BAB MNT;
RA   NARWA R., ROQUES P., COURPOTIN C., PARNETMATHIEU F., BOUSSIN F.,
RA   ROANE A., MARC D., LASFARGUES G., DORMONT D.;
RL   J. VIROL. 70:4474-4483(1996).
DR   EMBL; X99948; E258593; -.
KW   MATRIX PROTEIN.
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  14550 MW;  542AB924 CRC32;

  Query Match           3.0%;  Score 100;  DB 11;  Length 129;
  Best Local Similarity  41.9%;  Pred. No. 2.23e-01;
  Matches   18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db     28 KYR-LKHIVWASRELERFALNPGLLESAEGCQQLMEQL-QSTL 68
          ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   12
ID   O36786         PRELIMINARY;   PRT;   130 AA.
AC   O36786;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MA-P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LEIGH BROWN A.J., LOBIDEL D., WADE C.M., REBUS S., PHILLIPS N.,
RA   BRETTLE R.P., FRANCE A.J., LEEN C.S., MCMENAMIN J., MCMILLAN A.,
RA   MAW R.D., MULCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,
RA   PEUTHERER J.F.;
RL   VIROLOGY 235:166-177(1997).
DR   EMBL; AF014183; G2406708; -.
FT   NON_TER       1      1
FT   NON_TER     130    130
SQ   SEQUENCE   130 AA;  14584 MW;  428D9E75 CRC32;

  Query Match           3.0%;  Score 100;  DB 11;  Length 130;
  Best Local Similarity  41.9%;  Pred. No. 2.23e-01;
  Matches   18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db      5 KYK-LKHIVWASRELERFAVNPGLLETSEGCRQIMEQL-QPSL 45
          ||: |:: | |::| : |:: :|| |: : ||  || | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84
```

```
RESULT   13
ID   O36808         PRELIMINARY;   PRT;   130 AA.
AC   O36808;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MA-P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LEIGH BROWN A.J., LOBIDEL D., WADE C.M., REBUS S., PHILLIPS N.,
RA   BRETTLE R.P., FRANCE A.J., LEEN C.S., MCMENAMIN J., MCMILLAN A.,
RA   MAW R.D., MULCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,
     PEUTHERER J.F.;
     VIROLOGY 235:166-177(1997).
     EMBL; AF014205; G2406752; -.
FT   NON_TER       1      1
FT   NON_TER     130    130
SQ   SEQUENCE   130 AA;  14527 MW;  7325723C CRC32;

  Query Match             3.0%;  Score 100;  DB 11;  Length 130;
  Best Local Similarity  39.5%;  Pred. No. 2.23e-01;
  Matches    17;  Conservative   12;  Mismatches   9;  Indels   5;  Gaps   5;

Db       5 KYK-LKHIVWASRELERFAVNPGLLETSEGCRQIMEQL-QSAL 45
           ||: |:: | |::| : |:: :|| |: : ||  || | | :|
Qy      45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   14
ID   Q97725         PRELIMINARY;   PRT;   133 AA.
AC   Q97725;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE   MATRIX PROTEIN P17 (FRAGMENT).
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAL-M;
RA   NARWA R., ROQUES P., COURPOTIN C., PARNETMATHIEU F., BOUSSIN F.,
RA   ROANE A., MARC D., LASFARGUES G., DORMONT D.;
RL   J. VIROL. 70:4474-4483(1996).
DR   EMBL; Z79562; E261908; -.
KW   MATRIX PROTEIN.
FT   NON_TER     133    133
SQ   SEQUENCE   133 AA;  14724 MW;  E22E819C CRC32;

  Query Match             3.0%;  Score 100;  DB 11;  Length 133;
  Best Local Similarity  44.2%;  Pred. No. 2.23e-01;
  Matches    19;  Conservative    9;  Mismatches  10;  Indels   5;  Gaps   5;

Db      28 KYR-LKHIVWASRELERFALNPGLLEASEGCRQILGQL-QPSL 68
           ||| |:: | |::| : |:| :|| || : ||  :  | | ||
Qy      45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSL 84


RESULT   15
ID   Q77804         PRELIMINARY;   PRT;   486 AA.
AC   Q77804;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   GAG PROTEIN.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DJ258; TISSUE=BLOOD;
RA   LOUWAGIE J.J., MCCUTCHAN F., BRENNAN T., PEETERS M., BRENNAN T.,
RA   SANDERS-BUELL E., EDDY G., DER GROEN G., FRANSEN K.,
RA   GERSHY-DAMET M., DELEYS R., BURKE D.;
RL   SUBMITTED (JUN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L11763; G554842; -.
SQ   SEQUENCE   486 AA;  54094 MW;  6CB6088C CRC32;

  Query Match             3.0%;  Score 100;  DB 11;  Length 486;
  Best Local Similarity  32.9%;  Pred. No. 2.23e-01;
  Matches    24;  Conservative   18;  Mismatches  24;  Indels   7;  Gaps   7;

Db      28 KYR-LKHLVWASRELERFALNPGLLETAEGCQQLMEQL-QSAL-GTESK-ELKSLYNTIA 83
           ||| |:: | |::| : |:| :|| |: : |:  || | | :|     |  :|   | :
Qy      45 KYRTLRQCV-AGKETN-FSLASGL-EAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLR 101

Db      84 VLWCVHQRIDIKD 96
           : | : | :: :|
Qy     102 IYWSMYQSLQGND 114


Search completed: Tue Jun 23 18:23:21 1998
Job time : 137 secs.
```

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jun 24 20:41:51 1998; MasPar time 2627.46 Seconds
1357.834 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-3
Description: (1-2138) from US08866354.seq
Perfect Score: 2138
N.A. Sequence: 1 AGCTCGCTCTCCCGGGGCAG.....AGACAGGCCCGCAGCCGTCG 2138
Comp: TCGAGCGAGAGGGCCCCGTC.....TCTGTCCGGGCGTCGGCAGC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 457396 seqs, 834342348 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl54
1:em_ba 2:em_htg 3:em_hum1 4:em_hum2 5:em_in 6:em_om
7:em_or 8:em_ov 9:em_pat 10:em_pl 11:em_ro 12:em_vi
Database: genbank106
13:gb_ba 14:gb_htg 15:gb_in 16:gb_om 17:gb_ov 18:gb_pat
19:gb_ph 20:gb_pl 21:gb_pr1 22:gb_pr2 23:gb_ro 24:gb_st
25:gb_sy 26:gb_un 27:gb_vi

Statistics: Mean 11.685; Variance 5.512; scale 2.120

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	2138	100.0	2138	23	RNU59486	Rattus norvegicus GDNF	0.00e+00
2	2081	97.3	3616	23	RNU97142	Rattus norvegicus RET	0.00e+00
3	1615	75.5	2549	23	AF014117	Mus musculus GDNF rece	0.00e+00
4	1379	64.5	1651	23	AF015172	Mus musculus GDNF rece	0.00e+00
5	1302	60.9	1392	23	RN2072	Rattus norvegicus mRNA	0.00e+00
6	1279	59.8	1415	23	AB000800	Mouse mRNA for GDNF re	0.00e+00
7	1245	58.2	2560	22	AF038421	Homo sapiens GPI-linke	0.00e+00
8	1139	53.3	1619	22	AF042080	Homo sapiens glial cel	0.00e+00
9	1128	52.8	2175	22	HSU95847	Human GDNF receptor al	0.00e+00
10	1093	51.1	1707	22	HSU97144	Homo sapiens RET ligan	0.00e+00
11	640	29.9	777	23	AF012811	Mus musculus GDNF rece	0.00e+00
12	578	27.0	3037	17	GGU90541	Gallus gallus GDNF rec	0.00e+00
13	275	12.9	596	22	HSGFRA1G06	Homo sapiens GPI-linke	1.19e-205
14	244	11.4	497	22	HSGFRA1G03	Homo sapiens GPI-linke	2.41e-178
15	225	10.5	1395	22	HSU93703	Human glial cell line-	1.09e-161
16	225	10.5	1526	22	AF002700	Homo sapiens GDNF fami	1.09e-161
17	225	10.5	2906	22	HSU97145	Homo sapiens RET ligan	1.09e-161
18	220	10.3	1392	23	AF002701	Mus musculus GDNF fami	2.57e-157
19	215	10.1	1395	23	AF005226	Rattus norvegicus glia	5.98e-153
20	213	10.0	2787	23	RNU97143	Rattus norvegicus RET	3.32e-151
21	180	8.4	2933	17	GGU90542	Gallus gallus neurturi	1.47e-122
22	166	7.8	529	22	HSGFRA1G02	Homo sapiens GPI-linke	1.67e-110
23	145	6.8	432	22	HSGFRA1G09	Homo sapiens GPI-linke	1.46e-92
24	114	5.3	832	22	HSGFRA1G11	Homo sapiens GPI-linke	1.70e-66
25	112	5.2	385	18	A42902	Sequence 34 from Paten	7.76e-65
26	106	5.0	397	22	HSGFRA1G08	Homo sapiens GPI-linke	7.04e-60
27	98	4.6	436	22	HSGFRA1G07	Homo sapiens GPI-linke	2.55e-53
28	65	3.0	363	22	HSGFRA1G04	Homo sapiens GPI-linke	3.78e-27
c 29	59	2.8	7218	18	I66494	Sequence 14 from paten	1.29e-22
c 30	42	2.0	215	18	I28278	Sequence 5 from patent	1.97e-10
31	43	2.0	256	22	HSGFRA1G10	Homo sapiens GPI-linke	4.11e-11
32	42	2.0	7218	18	I66494	Sequence 14 from paten	1.97e-10
33	36	1.7	215	18	I28278	Sequence 5 from patent	1.79e-06
34	29	1.4	2602	23	S66545	putative alkaline phos	3.25e-02
c 35	31	1.4	129402	14	HS402G11	Human DNA sequence ***	2.21e-03
36	29	1.4	134226	27	IH1CG	Ictalurid herpesvirus	3.25e-02
c 37	28	1.3	390	23	S65019	mucin [rats, Sprague-D	1.20e-01
c 38	27	1.3	565	18	E04076	gDNA encoding envelope	4.32e-01
39	28	1.3	1203	22	AF051767	Homo sapiens GDNF fami	1.20e-01
40	28	1.3	10772	15	AF012089	Drosophila melanogaste	1.20e-01
c 41	28	1.3	134743	22	HUAC002300	Homo sapiens Chromosom	1.20e-01
c 42	28	1.3	143701	22	HS29C18	Homo sapiens DNA seque	1.20e-01
c 43	28	1.3	165197	22	HSAC002070	Human BAC clone 7E17 f	1.20e-01
44	27	1.3	171368	22	HSU95738	Human chromosome 16p13	4.32e-01
c 45	28	1.3	202233	22	AC003664	Homo sapiens chromosom	1.20e-01

ALIGNMENTS

RESULT 1
LOCUS RNU59486 2138 bp mRNA ROD 23-JUL-1996
DEFINITION Rattus norvegicus GDNF receptor alpha mRNA, complete cds.
ACCESSION U59486
NID g1399862
KEYWORDS .
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Rattus.
REFERENCE 1 (bases 1 to 2138)
AUTHORS Jing,S.J., Wen,D., Yu,Y., Holst,P.L., Luo,Y., Fang,M., Tamir,R., Antonio,L., Hu,Z., Cupples,R., Louis,J.-C., Hu,S., Altrock,B. and Fox,G.M.
TITLE GDNF-induced activation of the ret protein tyrosine kinase is mediated by GDNFR-alpha, a novel receptor for GDNF
JOURNAL Cell 85 (7), 1113-1124 (1996)
MEDLINE 96270513
REFERENCE 2 (bases 1 to 2138)
AUTHORS Fox,G.M., Jing,S.J., Yu,Y., Holst,P.L., Fang,M., Tamir,R., Antonio,L., Hu,Z. and Louis,J.-C.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1996) Immunology, Amgen, Inc., Amgen Center, Thousand Oaks, CA 91320-1789, USA
FEATURES Location/Qualifiers
source 1. .2138
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
CDS 302. .1708
/codon_start=1
/product="GDNF receptor alpha"
/db_xref="PID:g1399863"
/translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
IACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSN
CLKENYADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD

```
                    NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVL
                    PPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPV
                    LMLTALAALLSVSLAETS"
BASE COUNT      494 a    593 c    584 g    467 t
ORIGIN

  Query Match          100.0%;  Score 2138;  DB 23;  Length 2138;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  2138;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AGCTCGCTCTCCCGGGGCAGTGGTGTGGATGCACCGGAGTTCGGGCGCTGGGCAAGTTGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 AGCTCGCTCTCCCGGGGCAGTGGTGTGGATGCACCGGAGTTCGGGCGCTGGGCAAGTTGG 60

Db     61 GTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCCGGATCTGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCCGGATCTGA 120

      121 GTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGGATGGAGCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      121 GTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGGATGGAGCT 180

Db    181 GAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCTCTCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 GAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCTCTCTC 240

Db    241 CCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCGCGGCGGCAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 CCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCGCGGCGGCAC 300

Db    301 CATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 CATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGA 360

Db    361 GGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 GGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACA 420

Db    421 GAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 GAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTT 480

Db    481 CAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 CAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAA 540

Db    541 GCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 GCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCT 600

Db    601 GCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 GCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCC 660

Db    661 GTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 GTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGA 720

Db    721 TGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 TGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGC 780

Db    781 CTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 CTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCAC 840

Db    841 CAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 CAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTT 900

Db    901 CGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 CGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGC 960

Db    961 CTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 CTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAG 1020

Db   1021 GCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 GCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCT 1080

Db   1081 TGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 TGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGA 1140

Db   1141 GAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 GAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAA 1200

Db   1201 CTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 CTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAA 1260

Db   1261 TGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1261 TGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAA 1320

Db   1321 TGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1321 TGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGT 1380

Db   1381 CCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1381 CCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCC 1440

Db   1441 AGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1441 AGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGC 1500

Db   1501 TCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 TCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTT 1560

Db   1561 CGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 CGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1620

Db   1621 TCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTT 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1621 TCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTT 1680

Db   1681 ATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGACAAA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1681 ATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGACAAA 1740

Db   1741 AGAGAACCAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCT 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 AGAGAACCAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCT 1800

Db   1801 CCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAAAGCTTGTGG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801 CCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAAAGCTTGTGG 1860

Db   1861 CCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCATAAGGCTCTG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1861 CCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCATAAGGCTCTG 1920

Db   1921 GGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGGCTTATCATG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1921 GGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGGCTTATCATG 1980

Db   1981 TGTTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCTTGAACCCTG 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1981 TGTTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCTTGAACCCTG 2040
```

```
Db    2041 TTCTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCATGTCTTACTC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2041 TTCTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCATGTCTTACTC 2100

Db    2101 AGCAGCATTGCCTTCTGAAGACAGGCCCGCAGCCGTCG 2138
           |||||||||||||||||||||||||||||||||||||
Qy    2101 AGCAGCATTGCCTTCTGAAGACAGGCCCGCAGCCGTCG 2138


RESULT    2
LOCUS       RNU97142     3616 bp    mRNA            ROD       29-JUL-1997
DEFINITION  Rattus norvegicus RET ligand 1 (RETL1) mRNA, complete cds.
ACCESSION   U97142
NID         g2282021
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Rattus.
REFERENCE   1  (bases 1 to 3616)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Glial cell line-derived neurotrophic factor-dependent RET
            activation can be mediated by two different cell-surface accessory
            proteins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6238-6243 (1997)
  MEDLINE   97322356
REFERENCE   2  (bases 1 to 3616)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1997) Molecular Genetics, BIOGEN, 14 Cambridge
            Center, Cambridge, MA 02142, USA
FEATURES             Location/Qualifiers
     source          1. .3616
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /dev_stage="embryonic day 18"
                     /tissue_type="kidney"
     gene            1. .3616
                     /gene="RETL1"
     CDS             257. .1663
                     /gene="RETL1"
                     /note="PI-linked cell-surface accessory protein; GDNF
                     receptor alpha"
                     /codon_start=1
                     /product="RET ligand 1"
                     /db_xref="PID:g2282022"
                     /translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     IACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPV
                     LMLTALAALLSVSLAETS"
BASE COUNT      876 a    955 c    921 g    864 t
ORIGIN

  Query Match           97.3%;  Score 2081;  DB 23;  Length 3616;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches  2082;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      7 GCAGGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCC 66
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     52 GCAAGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCC 111

Db     67 CGGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTG 126
                                                                      

Qy    112 CGGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTG 171

Db    127 GATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTG 186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    172 GATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTG 231

Db    187 AGCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCG 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    232 AGCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCG 291

Db    247 CGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGAT 306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    292 CGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGAT 351

Db    307 GTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCT 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    352 GTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCT 411

Db    367 GAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGA 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    412 GAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGA 471

Db    427 AACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGA 486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    472 AACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGA 531

Db    487 GGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAA 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    532 GGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAA 591

Db    547 GAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGA 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    592 GAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGA 651

Db    607 AGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTT 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    652 AGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTT 711

Db    667 CATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGC 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    712 CATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGC 771

Db    727 AGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCC 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    772 AGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCC 831

Db    787 CTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAG 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    832 CTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAG 891

Db    847 GCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    892 GCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 951

Db    907 GGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGA 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    952 GGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGA 1011

Db    967 ACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAG 1026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1012 ACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAG 1071

Db   1027 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 1086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1072 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 1131

Db   1087 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 1146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1132 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 1191

Db   1147 GACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAA 1206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   1192 GACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAA 1251

Db   1207 CAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTG 1266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1252 CAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTG 1311

Db   1267 TCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGC 1326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1312 TCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGC 1371

Db   1327 CCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCC 1386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1372 CCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCC 1431

Db   1387 TCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAA 1446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1432 TCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAA 1491

     1447 TTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGA 1506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1492 TTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGA 1551

Db   1507 TAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAAT 1566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1552 TAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAAT 1611

Db   1567 GGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGC 1626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1612 GGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGC 1671

Db   1627 TGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGA 1686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1672 TGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGA 1731

Db   1687 AAAGACAAAAGAGAACCAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATCCAGTT 1746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1732 AAAGACAAAAGAGAACCAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATCCAGTT 1791

Db   1747 TTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAA 1806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1792 TTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAA 1851

Db   1807 AGCTTGTGGCCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCAT 1866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1852 AGCTTGTGGCCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCAT 1911

Db   1867 AAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGG 1926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1912 AAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGG 1971

Db   1927 CTTATCATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCT 1986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1972 CTTATCATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCT 2031

Db   1987 TGAACCCTGTTCTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCAT 2046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2032 TGAACCCTGTTCTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCAT 2091

Db   2047 GTCTTACTCAGCAGCATTGCCTTCTGAAGACAGGCCCGCAGCC 2089
          |||||||||||||||||||||||||||||||||||||||||||
Qy   2092 GTCTTACTCAGCAGCATTGCCTTCTGAAGACAGGCCCGCAGCC 2134


RESULT    3
LOCUS       AF014117     2549 bp    mRNA            ROD       20-NOV-1997
DEFINITION  Mus musculus GDNF receptor alpha (GDNFR-alpha) mRNA, complete cds.
ACCESSION   AF014117
NID         g2624960
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1  (bases 1 to 2549)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Cloning of a novel murine isoform of the glial cell line-derived
            neurotrophic factor receptor
  JOURNAL   Neuroreport 9 (1) (1998) In press
REFERENCE   2  (bases 1 to 2549)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1997) Biochemistry, 10 Kent Ridge Crescent,
            Singapore 119260, Singapore
FEATURES             Location/Qualifiers
     source          1. .2549
                     /organism="Mus musculus"
                     /strain="C57"
                     /db_xref="taxon:10090"
                     /tissue_type="liver"
     gene            1. .2549
                     /gene="GDNFR-alpha"
     CDS             914. .2320
                     /gene="GDNFR-alpha"
                     /note="glial cell-line derived neurotrophic growth factor
                     receptor"
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2624961"
                     /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     VACTERRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPV
                     MVFTALAALLSVSLAETS"
BASE COUNT      623 a    680 c    693 g    553 t
ORIGIN

  Query Match           75.5%;  Score 1615;  DB 23;  Length 2549;
  Best Local Similarity  94.2%;  Pred. No. 0.00e+00;
  Matches  1785;  Conservative    0;  Mismatches  98;  Indels  12;  Gaps   4;

Db    663 GGCAGGTTGGGTCGGAACTGAACCCCTAAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGC 722
          |||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Qy     51 GGCAAGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGC 110

Db    723 TCGGAGCTGAGTCCCTGGCGGCGGTGGGCGGCAGAGCAACGGGGAGTCTGCTCTCACCCT 782
          |||| ||||||| |||||||||||||||||||| |||||||||||||||||||||||||
Qy    111 CCGGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCT 170

Db    783 GGATGGAGCTTAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCTCTGCACGCT 842
          |||||||||| |||||||||||||||||||||||||||||||||||||| |||||||||
Qy    171 GGATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCT 230

Db    843 GAGCTCTCTCCGCGAGATCCGGTGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGC 902
          |||||||||| ||||| | || ||||||||||||||||||||||||||||||||||||||
Qy    231 GAGCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGC 290

Db    903 GCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTGA 962
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Qy    291 GCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGA 350

Db    963 TGTCGGCCGAGGTGAGTGGTGGGGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGCC 1022
          |||| |||||||||||||||| ||||| |||||||||||||||||| ||||||||||
Qy    351 TGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCC 410

Db   1023 TGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAGG 1082
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Qy    411 TGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGG 470

Db   1083 AAACCAACTTCAGCCTGACATCCGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATGG 1142
          |||||||||||||||||||||||||||||| |||||||||||||||||| ||||| ||||
Qy    471 AAACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGG 530
```

```
Db   1143 AGGCCTTGAAGCAGAAGTCTCTCTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGA 1202
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Qy    531 AGGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGA 590

Db   1203 AGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTGG 1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Qy    591 AGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGG 650

Db   1263 AAGATTCCCCATACGAGCCGGTTAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCGT 1322
          |||||||||| || |||||||||||||||||| ||||||||||||||||||||||||||||
Qy    651 AAGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGT 710

Db   1323 TCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTCGATG 1382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| || |
Qy    711 TCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACG 770

Db   1383 CAGCCAAGGCCTGCAACCTGGATGACACCTGCAAGAAGTACAGGTCCGCCTACATCACCC 1442
          |||||||||||||||||||| |||||||| |||||||||||||| ||||||||||||||
Qy    771 CAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCC 830

Db   1443 CCTGTACCACCAGCATGTCCAATGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTCA 1502
          |||| |||||||||||||||| || |||||||||||||| |||||||||| |||||||
Qy    831 CCTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCA 890

Db   1503 GGCAGTTCTTCGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCC 1562
          ||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Qy    891 GGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCC 950

Db   1563 GGGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAG 1622
          ||||| |||||||||||||| ||||||||||||||||||||| ||||||||||||||||
Qy    951 GGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAG 1010

Db   1623 AACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCA 1682
          |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Qy   1011 AACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCA 1070

Db   1683 GATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACT 1742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1071 GATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACT 1130

Db   1743 GTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCA 1802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1131 GTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCA 1190

Db   1803 TGACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCA 1862
          ||||||| |||||| |||||||||||||||||| ||||| || ||||| || |||||||
Qy   1191 TGACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCA 1250

Db   1863 ACAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGT 1922
          |||| |||||||||||||||| ||| |||| ||||||||||||||||||||||||||| |
Qy   1251 ACAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTTAAGGACAATACTT 1310

Db   1923 GTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAACCAG 1982
          |||||||||||||||||||||||||||||||||| |||||||||||||||| ||||
Qy   1311 GTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAG 1370

Db   1983 CCCCCCCAGTCCAGACCACCACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGC 2042
          |||| |||||||||||||||||||| |||||||||||||||| ||||||||||||
Qy   1371 CCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGC 1430

Db   2043 CTCTAGGGCCAGCAGGCTCTGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTA 2102
          |||| |||||||||||| |||||||||||||| |||||||||||||||||||| |||| |
Qy   1431 CTCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGA 1490

Db   2103 ATTTGCAGGCACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTG 2162
          |||||||||| |||||||||||||||||||| ||||| || ||||| ||||||||||||
Qy   1491 ATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTG 1550

Db   2163 ATAATGATTACGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAA 2222
          ||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1551 ATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAA 1610

Db   2223 TGGCTGCTCCTCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGG 2282
          |||||||||||||||||||| |||||||||||||||||| ||| || ||||||| || |
Qy   1611 TGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTG 1670

Db   2283 CTGCCCTGTTGTCTGTATCATTGGCAGAAACATCGTAGCTGCATCC--------AGTATG 2334
          |||||||||| |||||||| |||||||||| |||||||||||||         ||||||
Qy   1671 CTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATG 1730

Db   2335 AAAAGACAAAAGAAAACCAAGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAG 2394
          |||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||
Qy   1731 AAAAGACAAAAGAGAACCAAGTATTCTGTCCC-TGTCCTCTTGTATATCTGAAAATCCAG 1789

Db   2395 TTTTAAAAGCTCGGTTGAGAAGCAGTTTCACCCAGCGGGAACTCCTTTGTTGTTTTTAAG 2454
          |||||||||||| |||||||||||||||||||| | ||||||| ||  ||||||||||
Qy   1790 TTTTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAG 1849

Db   2455 AAAGCTTTTTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAA 2514
          ||||||| | |  |||||||  ||||||||||||||||||||||||||||||||||||
Qy   1850 AAAGCTTGT-G--GCCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAA 1906

Db   2515 CTCAGAAAACTCTGGGGCATGGTGTGGCTTCAGGG 2549
          | || ||  |||||||| ||||| ||||| ||||
Qy   1907 CCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGG 1941


RESULT    4
LOCUS       AF015172     1651 bp    mRNA            ROD       20-NOV-1997
DEFINITION  Mus musculus GDNF receptor beta (GDNFR-beta) mRNA, complete cds.
ACCESSION   AF015172
NID         g2624962
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1651)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Cloning of a novel murine isoform of the glial cell line-derived
            neurotrophic factor receptor
  JOURNAL   Neuroreport 9 (1) (1998) In press
REFERENCE   2  (bases 1 to 1651)
  AUTHORS   Dey,B.K., Wong,Y.W. and Too,H.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1997) Biochemistry, National University of
            Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore
FEATURES             Location/Qualifiers
     source          1. .1651
                     /organism="Mus musculus"
                     /strain="C57"
                     /db_xref="taxon:10090"
     gene            1. .1651
                     /gene="GDNFR-beta"
     CDS             230. .1621
                     /gene="GDNFR-beta"
                     /note="glial cell-line derived neurotrophic factor
                     receptor beta."
                     /codon_start=1
                     /product="GDNF receptor beta"
                     /db_xref="PID:g2624963"
                     /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISVEHISKGNNCLDAAKACNLDD
                     TCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTE
                     RRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKEN
                     YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTA
                     LAALLSVSLAETS"
BASE COUNT      394 a    475 c    443 g    339 t
ORIGIN

  Query Match           64.5%;  Score 1379;  DB 23;  Length 1651;
  Best Local Similarity  94.2%;  Pred. No. 0.00e+00;
  Matches  1549;  Conservative    0;  Mismatches  80;  Indels  15;  Gaps   1;
```

```
Db      1 CCCCTAAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCTCGGAGCTGAGTCCCTGGCGGC 60
          ||||| ||||||||||||||||||||||||||||||||| |||| ||||||| ||||||||
Qy     73 CCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCCGGATCTGAGTCGCTGGCGGC 132

Db     61 GGTGGGCGGCAGAGCAACGGGGAGTCTGCTCTCACCCTGGATGGAGCTTAACTTTGAGTG 120
          |||||||||||||||| |||||||||||||||||||||||||||||||| ||||||||||
Qy    133 GGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGGATGGAGCTGAACTTTGAGTG 192

Db    121 GCCAGAGGAGCGCAGTCGCCCGGGGATCTCTGCACGCTGAGCTCTCTCCGCGAGATCCGG 180
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| | ||
Qy    193 GCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCTCTCTCCCCGAGACCGGG 252

Db    181 TGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCGCGGCGGCACCATGTTCCTAGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    253 CGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCGCGGCGGCACCATGTTCCTAGC 312

Db    241 CACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTGATGTCGGCCGAGGTGAGTGGTGG 300
          |||||||||||||| |||||||||||||||||||||||| |||||||||||||||||||
      313 CACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAGGTGAGTGGTGG 372

Db    301 GGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGCCTGAAGGAACAGAGCTGCAGCAC 360
          |||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Qy    373 AGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAGAGCTGCAGCAC 432

Db    361 CAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTGACATC 420
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Qy    433 CAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTGACATC 492

Db    421 CGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATGGAGGCCTTGAAGCAGAAGTCTCT 480
          |||||| |||||||||||||||||||| ||||| ||||||||||||||||||||||||||
Qy    493 CGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAGCAGAAGTCTCT 552

Db    481 CTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATCTACTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    553 GTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATCTACTG 612

Db    541 GAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTGGAAGATTCCCCATACGAGCCGGT 600
          |||||||||||||||||||||||||||||||||| |||||||||||||| || ||||||||
Qy    613 GAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCGTATGAGCCGGT 672

Db    601 TAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAG---------- 647
          |||||||||| |||||||||||||||||||||||||||||||||||
Qy    673 TAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTTCCAGCA 732

Db    648 --TGGAACACATTTCCAAAGGGAACAACTGCCTCGATGCAGCCAAGGCCTGCAACCTGGA 705
            |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Qy    733 AGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGA 792

Db    706 TGACACCTGCAAGAAGTACAGGTCCGCCTACATCACCCCCTGTACCACCAGCATGTCCAA 765
          |||||||| |||||||||||||| |||||||||||||||| ||||||||||||||||||
Qy    793 CGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACCAGCATGTCCAA 852

Db    766 TGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTCAGGCAGTTCTTCGACAAAGTTCC 825
          || |||||||||||||| |||||||||| ||||||||||||||||||||| |||||
Qy    853 CGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCC 912

Db    826 AGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACGTCGCCTGCACCGAGAG 885
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||| |
Qy    913 GGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCG 972

Db    886 GCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCT 945
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Qy    973 GCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCT 1032

Db    946 GAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATTTTTT 1005
          || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy   1033 GAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATTTTTT 1092

Db   1006 TACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGA 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1093 TACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGA 1152

Db   1066 CTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAGACTC 1125
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||||
Qy   1153 CTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAGACTC 1212

Db   1126 CAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGGAAGA 1185
          |||||||||||| ||||| || ||||| || |||||||||| ||||||||||||||||||
Qy   1213 CAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGGAAGA 1272

Db   1186 TTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTCAAGC 1245
          ||| |||| |||||||||||||||||||||||| ||||||||||||||||||||||||||
Qy   1273 CTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTCAAGC 1332

Db   1246 CTTTGGCAATGGCTCGGATGTGACCATGTGGCAACCAGCCCCCCCAGTCCAGACCACCAC 1305
          |||||||||||||| |||||||||||||||| |||||||| ||||||||||||||||||
Qy   1333 CTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCACCAC 1392

Db   1306 TGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCTCTGA 1365
          |||||| |||||||||||||||||| |||||||||||||||| |||||||||| |||||
Qy   1393 TGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGTCTGA 1452

Db   1366 GAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGCACAGAAGCTGAA 1425
          |||||||||| |||||||||||||||||||| ||||| |||||||||| ||||||||||
Qy   1453 GAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAAGCTGAA 1512

Db   1426 ATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTACGGAAAGGATGG 1485
          |||||||||| ||||| || ||||| |||||||||||||||| ||||| ||||||||||
Qy   1513 ATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGGAAAGGATGG 1572

Db   1486 TCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTGCGG 1545
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Qy   1573 TCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTGCAG 1632

Db   1546 TCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTTGTCTGTATCATT 1605
          |||||||||||||||||| || || ||||||| || |||||||||| |||||||| ||
Qy   1633 TCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTATCTGTATCGTT 1692

Db   1606 GGCAGAAACATCGTAGCTGCATCC 1629
          ||||||||| ||||||||||||||
Qy   1693 GGCAGAAACGTCGTAGCTGCATCC 1716
```

```
RESULT    5
LOCUS       RN2072       1392 bp    RNA             ROD       24-OCT-1997
DEFINITION  Rattus norvegicus mRNA for GDNFR-alpha/TrnRl-delta protein,
            complete cds.
ACCESSION   AJ002072
NID         g2564301
KEYWORDS    GDNFR-alpha/TrnRl-delta protein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Zhong,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1997) Zhong J., Molecular Neurobiochemistry
            NC7/172, Ruhr University Bochum, Universitaetsstr. 150, D-44780
            Bochum, GERMANY
REFERENCE   2  (bases 1 to 1392)
  AUTHORS   Zhong,J., Annies,M. and Heumann,R.
  TITLE     Rattus norvegicus GDNFR-alpha/TrnRl-delta mRNA, complete cds
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Rattus norvegicus"
                     /strain="Hannover"
                     /db_xref="taxon:10116"
     CDS             1. .1392
                     /codon_start=1
                     /product="GDNFR-alpha/TrnRl-delta protein"
                     /db_xref="PID:e1154274"
                     /db_xref="PID:g2564302"
                     /translation="MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
```

```
                  KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                  YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISVEHISKGNNCLDAAKACNLDD
                  TCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                  RRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKEN
                  YADCLLAYSGLIGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLK
                  NAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCAN
                  LQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTA
                  LAALLSVSLAETS"
BASE COUNT      338 a    406 c    356 g    292 t
ORIGIN

  Query Match           60.9%;  Score 1302;  DB 23;  Length 1392;
  Best Local Similarity  98.9%;  Pred. No. 0.00e+00;
  Matches  1392;  Conservative    0;  Mismatches   0;  Indels  15;  Gaps   1;

Db      1 ATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    302 ATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAG 361

Db     61 GTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    362 GTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAG 421

Db    121 AGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    422 AGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTC 481

Db    181 AGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    482 AGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAG 541

Db    241 CAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    542 CAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTG 601

Db    301 CGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    602 CGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCG 661

Db    361 TATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAG-- 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    662 TATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGAT 721

Db    419 ------------TGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCC 465
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    722 GTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCC 781

Db    466 TGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACC 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    782 TGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACC 841

Db    526 AGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTC 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    842 AGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTC 901

Db    586 GACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCC 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    902 GACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCC 961

Db    646 TGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGG 705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    962 TGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGG 1021

Db    706 CCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTT 765
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1022 CCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTT 1081

Db    766 GCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAG 825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1082 GCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAG 1141

Db    826 AACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAAC 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1142 AACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAAC 1201

Db     886 TACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAAT 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1202 TACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAAT 1261

Db     946 GACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAAT 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1262 GACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAAT 1321

Db    1006 GCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTC 1065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1322 GCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTC 1381

Db    1066 CAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCA 1125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1382 CAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCA 1441

Db    1126 GCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCT 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1442 GCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCT 1501

Db    1186 CAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTC 1245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1502 CAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTC 1561

Db    1246 GGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCT 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1562 GGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCT 1621

Db    1306 CCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTA 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1622 CCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTA 1681

Db    1366 TCTGTATCGTTGGCAGAAACGTCGTAG 1392
           ||||||||||||||||||||||||||||
Qy   1682 TCTGTATCGTTGGCAGAAACGTCGTAG 1708


RESULT    6
LOCUS       AB000800     1415 bp    mRNA            ROD       04-FEB-1997
DEFINITION  Mouse mRNA for GDNF receptor alpha, complete cds.
ACCESSION   AB000800
NID         g1816441
KEYWORDS    GDNF receptor alpha.
SOURCE      Mus musculus Dorsal root ganglion Schwann cell cell_line:IMS32 cDNA
            to mRNA.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 1415)
  AUTHORS   Watabe,K.
  TITLE     Regulation of GDNF receptor alpha mRNA in Schwann cell cultures
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 1415)
  AUTHORS   Watabe,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1997) to the DDBJ/EMBL/GenBank databases.
            Kazuhiko Watabe, Tokyo Metropolitan Institute for Neuroscience,
            Neuropathology; Musashidai 2-6, Fuchu, Tokyo 183, Japan
            (E-mail:kazwtb@tmin.ac.jp, Tel:0423-25-3881, Fax:0423-21-8678)
FEATURES             Location/Qualifiers
     source          1. .1415
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="IMS32"
                     /cell_type="Schwann cell"
                     /tissue_type="Dorsal root ganglion"
     primer_bind     1. .20
     CDS             2. .1408
                     /standard_name="glial cell line-derived neurotrophic
                     factor receptor alpha"
```

```
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:d1019934"
                     /db_xref="PID:g1816442"
                     /translation="MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKA
                     CNLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     VACTERRRQTIVPVCSYEERERPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSN
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTMWQPAPPVQTTTAMTTTAFRIKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPV
                     MVFTALAALLSVSLAETS"
     primer_bind     complement(1396. .1415)
BASE COUNT      351 a    402 c    358 g    304 t
ORIGIN
```

Query Match 59.8%; Score 1279; DB 23; Length 1415;
st Local Similarity 95.2%; Pred. No. 0.00e+00;
tches 1347; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Db      1 CATGTTCCTAGCCACTCTGTACTTCGTGCTGCCACTCCTGGATTTGCTGATGTCGGCCGA 60
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||| |||||
Qy    301 CATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGA 360

Db     61 GGTGAGTGGTGGGGACCGCCTGGACTGTGTGAAAGCCAGTGATCAGTGCCTGAAGGAACA 120
          |||||||||||| ||||| |||||||||||||||||| ||||||||||||||||||||
Qy    361 GGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACA 420

Db    121 GAGCTGCAGCACCAAGTACCGCACACTGAGGCAGTGCGTGGCGGGCAAGGAAACCAACTT 180
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Qy    421 GAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTT 480

Db    181 CAGCCTGACATCCGGCCTCGAGGCCAAGGATGAGTGCCGCAGCGCTATGGAGGCCTTGAA 240
          |||||||||||||||| ||||||||||||||||||| ||||| ||||||||||||||||
Qy    481 CAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAA 540

Db    241 GCAGAAGTCTCTCTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCT 300
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 GCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCT 600

Db    301 GCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTACTGGAAGATTCCCC 360
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Qy    601 GCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCC 660

Db    361 ATACGAGCCGGTTAACAGCAGGCTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGA 420
          || |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy    661 GTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGA 720

Db    421 TGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTCGATGCAGCCAAGGC 480
          |||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Qy    721 TGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGC 780

Db    481 CTGCAACCTGGATGACACCTGCAAGAAGTACAGGTCCGCCTACATCACCCCCTGTACCAC 540
          |||||||||||| |||||||| |||||||||||||| |||||||||||||||| |||||
Qy    781 CTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCAC 840

Db    541 CAGCATGTCCAATGAAGTCTGCAACCGCCGCAAGTGCCACAAAGCCCTCAGGCAGTTCTT 600
          |||||||||||| || |||||||||||| |||||||||| ||||||||||||||||||||
Qy    841 CAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTT 900

Db    601 CGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACGTCGC 660
          |||||| ||||| |||||||||||||||||||||||||||||||||||||||||| ||||
Qy    901 CGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGC 960

Db    661 CTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAGAACGAGAGAG 720
          |||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Qy    961 CTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAG 1020

Db    721 GCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCT 780
          |||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||||
Qy   1021 GCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCT 1080

Db    781 TGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 TGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGA 1140

Db    841 GAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAA 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Qy   1141 GAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAA 1200

Db    901 CTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAA 960
          |||| |||||||||||||||||| ||||| || ||||| || |||||||||||| |||||
Qy   1201 CTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAA 1260

Db    961 TGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAA 1020
          |||||||||||| ||| |||| ||||||||||||||||||||||||||| ||||||||||
Qy   1261 TGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAA 1320

Db   1021 TGCAATTCAAGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCAGT 1080
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||
Qy   1321 TGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGT 1380

Db   1081 CCAGACCACCACTGCCATGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCC 1140
          |||||||||||||||  |||||||||||||||||| |||||||||||||||| |||||
Qy   1381 CCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCC 1440

Db   1141 AGCAGGCTCTGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGC 1200
          |||||| |||||||||||||| |||||||||||||||||||| ||||| ||||||||||
Qy   1441 AGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGC 1500

Db   1201 ACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTA 1260
            |||||||||||||||||| ||||| || ||||| |||||||||||||||| |||||
Qy   1501 TCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTT 1560

Db   1261 CGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561 CGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCC 1620

Db   1321 TCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTT 1380
          |||||||||| |||||||||||||||||| || || ||||||| || ||||||||||||
Qy   1621 TCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTT 1680

Db   1381 GTCTGTATCATTGGCAGAAACGTCGTAGCTGCATC 1415
          |||||||| ||||||||||||||||||||||||||
Qy   1681 ATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATC 1715
```

```
RESULT    7
LOCUS       AF038421     2560 bp    mRNA            PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) mRNA, complete cds.
ACCESSION   AF038421
NID         g2921531
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2560)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 2560)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .2560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
```

```
    gene            <1. .>2560
                    /gene="GFRA1"
    CDS             557. .1954
                    /gene="GFRA1"
                    /note="similar to GDNF Receptor alpha; part of functional
                    ligand for RET receptor tyrosine kinase"
                    /codon_start=1
                    /product="GPI-linked anchor protein"
                    /db_xref="PID:g2921532"
                    /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                    KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                    YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKA
                    CNLDDICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                    IACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSS
                    CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKD
                    NTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVL
                    PPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVL
                    VVTALSTLLSLTETS"
BASE COUNT      642 a    675 c    632 g    611 t
ORIGIN

Query Match           58.2%;  Score 1245;  DB 22;  Length 2560;
Best Local Similarity  84.9%;  Pred. No. 0.00e+00;
Matches  1625;  Conservative    0;  Mismatches 272;  Indels  18;  Gaps  12;

Db    308 CAGGTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGC--CATCCC 365
          || ||||||||||| |||||||||| |||||||  |||||||||||||||||  |  |||
Qy     53 CAAGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCC 112

Db    366 GGAGCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGG 424
          ||| |||||||||| |||||||||| | || |   |||  ||||| | ||||  | ||||
Qy    113 GGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGG 172

Db    425 ATGGAGCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGA 484
          |||||||||||||||| | |||||||| ||| |||  | ||||||||||||||||||||
Qy    173 ATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGA 232

Db    485 GCTCCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCC 544
          |||| |||  | |||||  |||||||||  |||||||   ||||||||||||||||||
Qy    233 GCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTG 289

Db    545 CGCGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTC 604
          |||| ||||||||||||||| || || |||||||||||||||||| ||| |||| |||||
Qy    290 CGCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTG 349

Db    605 CTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGC 664
          |||| ||||| ||||| || |||||||| ||||| || |||||||||| |||||||||
Qy    350 ATGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGC 409

Db    665 CTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAG 724
          |||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||
Qy    410 CTGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAG 469

Db    725 GAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATG 784
          || |||||||||||||| |||||||||| |||||||||||||||||||| |||||||||
Qy    470 GAAACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATG 529

Db    785 GAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAG 844
          |||||| |||||||||||| || ||||||||||||||||||||||||| |||||||| |||
Qy    530 GAGGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAG 589

Db    845 AAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTG 904
          ||||| || ||||| || ||||||||||||||||||||||||||||||||||||| || |||
Qy    590 AAGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTG 649

Db    905 GAGGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCA 964
          || |||||||| ||||| || |||||||||| ||||||||||||||||||  |||||
Qy    650 GAAGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCG 709

Db    965 TTCATATCAGATGTTTTTCAGCAAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGAT 1024
          |||||||||||||||| |||||||||| |||||| ||||||||||||||||||||||||
Qy    710 TTCATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGAC 769

Db   1025 GCAGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACC 1084
          ||||| |||||||||||||| |||||||  || ||||||||||||||||| |||||||||
Qy    770 GCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACC 829

Db   1085 CCGTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTC 1144
          || |||||||||||| |||||||||| |||||||||||||| ||||||||||||||||||
Qy    830 CCCTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTC 889

Db   1145 CGGCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGC 1204
           |||||||||| |||||||| |||||||||||||||||||| ||||||||||||||||||
Qy    890 AGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGC 949

Db   1205 CGGGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAA 1264
          |||||||||||||||| ||||| |||||||||| ||||| || ||||||||||||||||
Qy    950 CGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAA 1009

Db   1265 GAGAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGC 1324
          ||  | |||| |||||||||  ||| | |||| |||||||||||||| ||||||||||||
Qy   1010 GAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGC 1069

Db   1325 AGATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGC 1384
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |
Qy   1070 AGATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAAC 1129

Db   1385 TGTCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTC 1444
          ||||| ||||| |||||||| |||||||||| |||||||||| || ||||||||||||
Qy   1130 TGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTC 1189

Db   1445 ATGACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGC 1504
          ||||| |||||||||| |||||||||| |||||||| ||||| ||||||||||||||||
Qy   1190 ATGACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGC 1249

Db   1505 AACAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACA 1564
          ||||| || || ||||| ||||| |||||||||||| ||||||| || ||||||||||
Qy   1250 AACAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACT 1309

Db   1565 TGTCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCA 1624
          ||||| |||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Qy   1310 TGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCA 1369

Db   1625 GCCTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAG 1684
          |||   ||||| |||||||||||||||| || ||||||||| ||||||| |||||||||
Qy   1370 GCCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAG 1429

Db   1685 CCCCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCA 1744
          || |||||||||||||||||||||||||| || ||||| || ||||| ||||| |||||
Qy   1430 CCTCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCG 1489

Db   1745 AATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCC 1804
          ||||| ||||| |||||||||||||||||||||||||||||| |  ||||||||||| ||||
Qy   1490 AATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCT 1549

Db   1805 AATGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCA 1861
          || || |||  | ||| || |||||||   |||| ||||||||||||||||||||||||
Qy   1550 GATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCA 1609

Db   1862 ATGGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTG 1921
          |||||||||||||| |||||  |||||||| |||||| ||| ||| || | ||||| ||
Qy   1610 ATGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTT 1669

Db   1922 TCCACCCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACA 1981
          |  |||| |||||| || |     ||  |  |  | |  |      ||  |  |
Qy   1670 GCTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTAT 1729

Db   1982 TGTAAAAAGACAAA-AACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCA 2040
            || |  | | | ||||||||  |||||  ||||| |||||||||| |||||| ||||
Qy   1730 GAAAAGACAAAAGAGAACCAAGTATTCTGTC-CCTGTCCTCTTGTATATCTGAAAATCCA 1788

Db   2041 GTTT-AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTCCTTT 2099
          |||| |  ||||| |||||||| |||||  ||  |||||||||| | |||  || |  |||
Qy   1789 GTTTTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAAC-TCTTTCCTTGT--TTT 1845

Db   2100 TAAGAAAGCTTCTTGTGATCCTTCGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATC 2159
```

```
          |||||||||||| | | |  |||  ||||||||||  || | ||| | ||| |||   ||
Qy   1846 TAAGAAAGCTTGT-G-GC-CCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTC 1902

Db   2160 CAAACTCAGAAGGCTTTGGGATATGCTGTATTTTAAAGGGACAGTTTGTAACTTG 2214
          ||||| || |||||| ||||   || ||    |||| |||||  |||| | | ||
Qy   1903 CAAACCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATG 1957


RESULT    8
LOCUS       AF042080     1619 bp    mRNA            PRI       22-JAN-1998
DEFINITION  Homo sapiens glial cell line-derived neurotrophic factor receptor
            alpha (GFRA1) mRNA, complete cds.
ACCESSION   AF042080
NID         g2801556
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1619)
  AUTHORS   Shefelbine,S.E., Khorana,S., Schultz,P.N., Huang,E., Thobe,N.,
            Hu,Z.J., Fox,G.M., Jing,S., Cote,G.J. and Gagel,R.F.
  TITLE     Mutational analysis of the GDNF/RET-GDNFRa signaling complex in a
            kindred with vesicoureteral reflux
  JOURNAL   Hum. Genet. (1998) In press
REFERENCE   2  (bases 1 to 1619)
  AUTHORS   Shefelbine,S.E., Khorana,S., Schultz,P.N., Huang,E., Thobe,N.,
            Hu,Z.J., Fox,G.M., Jing,S., Cote,G.J. and Gagel,R.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1998) Endocrinology-Box 15, M.D. Anderson Cancer
            Center, 1515 Holcombe Blvd, Houston, TX 77030, USA
FEATURES             Location/Qualifiers
     source          1. .1619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25-q26"
                     /tissue_type="medullary thyroid carcinoma tumor"
     gene            1. .1619
                     /note="GDNFRa"
                     /gene="GFRA1"
     CDS             111. .1508
                     /gene="GFRA1"
                     /note="GPI-linked receptor"
                     /codon_start=1
                     /product="glial cell line-derived neurotrophic factor
                     receptor alpha"
                     /db_xref="PID:g2801557"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKA
                     CNLDDICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRD
                     IACTERRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSS
                     CLKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKD
                     NTCLKNAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVL
                     PPCANLQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVL
                     VVTALSTLLSLTETS"
     misc_feature    529. .543
                     /gene="GFRA1"
                     /note="found to be alternative exon after partial analysis
                     of genomic sequence"
BASE COUNT      409 a    448 c    419 g    343 t
ORIGIN

  Query Match           53.3%;  Score 1139;  DB 22;  Length 1619;
  Best Local Similarity  86.5%;  Pred. No. 0.00e+00;
  Matches  1404;  Conservative    0;  Mismatches 211;  Indels   9;  Gaps   5;

Db      1 CAGAGCAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCTCCCTCGGCAAGACCCAGCG 60
          ||||| ||| |||  | |||||||||||||||||||||||| |||  | |||||  |||
Qy    195 CAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCTCTCTCCCCGAGACCGGGCG 254

Db     61 GCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGCCGGCACCATGTTCCTGG 120
          |||||| |||||||  |||||||||||||||||||||||||||| |||| ||||||||||||||| |
Qy    255 GCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTGCGCGGCGGCACCATGTTCCTAG 311

Db    121 CGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTCGGCCGAAGTGAGCGGCG 180
          | || |||||||||||||||| ||| |||| |||||  |||| ||||| ||||| || |
Qy    312 CCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAGGTGAGTGGTG 371

Db    181 GAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAAGGAGCAGAGCTGCAGCA 240
          ||||||| ||||| || |||||||||| |||||||||||||||| ||||||||||||
Qy    372 GAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAGAGCTGCAGCA 431

Db    241 CCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGACCAACTTCAGCCTGGCAT 300
          |||||||||||||| |||||||||||||||||||||||||||||| |||||||||||||| |||
Qy    432 CCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTGACAT 491

Db    301 CCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGCCCTGAAGCAGAAGTCGC 360
          ||||||| |||||||||||||||||||| |||||||||||||| |||||||||||||| |
Qy    492 CCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAGCAGAAGTCTC 551

Db    361 TCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAACTGCCTGCGCATTTACT 420
          | |||||||||||||||||||||| |||||||| |||||||| || ||||| || ||||
Qy    552 TGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATCTACT 611

Db    421 GGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGATTCCCCATATGAACCAG 480
          |||||||||||||||||||||||||||||| || ||||| |||||||| ||||| || |
Qy    612 GGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCGTATGAGCCGG 671

Db    481 TTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCATATCAGATGTTTTTCAGC 540
          |||||||||| |||||||||||||||||||| ||||| |||||||||||||||| ||||
Qy    672 TTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTTCCAGC 731

Db    541 AAGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCG 600
          ||||||| |||||| |||||||||||||||||||||||| ||||| |||||||||||||| |
Qy    732 AAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGG 791

Db    601 ACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCA 660
          ||||||  || |||||||||||||||| |||||||||| |||||||||||| ||||||
Qy    792 ACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACCAGCATGTCCA 851

Db    661 ACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCC 720
          |||| |||||||||||||| |||||||||||||||| |||||||||| |||||||| |
Qy    852 ACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTC 911

Db    721 CGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGC 780
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||| ||||
Qy    912 CGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGC 971

Db    781 GGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTT 840
          || |||||||||| ||||| || |||||||||||||||||  | |||| |||||||||
Qy    972 GGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCC 1031

Db    841 TGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATCTCGCCTTGCGGATTTTT 900
          ||| | |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||
Qy   1032 TGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATTTTT 1091

Db    901 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCTAAAGGAAAACTACGCTG 960
          |||||||||||||||||||||||||||||||||||||| |||||| ||||| |||||||| |
Qy   1092 TTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAG 1151

Db    961 ACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGACCCCCAACTACATAGACT 1020
          |||||||||| |||||||||| || |||||||||||||||| |||||||| ||||||
Qy   1152 ACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAGACT 1211

Db   1021 CCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAGTGGGAACGACCTAGAAG 1080
          |||| |||||||| |||| |||||||||||||||||||||||| || || |||| ||||
Qy   1212 CCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGGAAG 1271

Db   1081 AGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATGTCTTAAAAATGCAATTCAAG 1140
          | |||||||||||| ||||||| || |||||||||| |||| ||||||||||||||||
Qy   1272 ACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTCAAG 1331

Db   1141 CCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTTCCCAGTACAGACCACCA 1200
          |||||||||||||||| ||||||||| |||||||||||||   |||| ||||||||||
```

```
Qy   1332 CCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCACCA 1391

Db   1201 CTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCCCCTGGGGCCAGCAGGGTCTG 1260
          ||||||| || |||||||||| ||||||| |||||||||| ||||||||||||||||||
Qy   1392 CTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGTCTG 1451

Db   1261 AGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTTACAGGCACAGAAGCTGA 1320
          ||||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||||||||
Qy   1452 AGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAAGCTGA 1511

Db   1321 AATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGGTAATTATGAAAAAGAAG 1380
          |||||||||||||||| |  |||||||||||| |||| || || |||  | ||| || |
Qy   1512 AATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGGAAAGGATG 1571

Db   1381 GTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCAAGCTGTG 1437
          ||||||   |||| |||||||||||||||||||||||||||||||||||||| |||||
Qy   1572 GTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTGCA 1631

Db   1438 GTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCACCCTATTATCTTTAACAG 1497
          |||||||| |||||| ||| ||| || | ||||| ||  |  |||| |||||| || |
     1632 GTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTATCTGTATCGT 1691

     1498 AAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAAAAAGACAAA-AACCAAG 1556
             ||  |  |  | |      |      || |  |      || |  | | | |||||||
Qy   1692 TGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGACAAAAGAGAACCAAG 1751

Db   1557 TTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTT-AGGAGCTCAGTTGAGAA 1615
          |  |||||  ||||| |||||||||| |||||| |||||||| |  ||||| |||||||||
Qy   1752 TATTCTGTC-CCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCTCCGTTGAGAA 1810

Db   1616 ACAG 1619
           |||
Qy   1811 GCAG 1814
```

```
RESULT    9
LOCUS       HSU95847     2175 bp    mRNA            PRI       02-OCT-1997
DEFINITION  Human GDNF receptor alpha mRNA, complete cds.
ACCESSION   U95847
NID         g2459741
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2175)
  AUTHORS   Hishiki,T., Kondoh,K., Ichimiya,S., Nimura,Y., Seki,N., Ozaki,T.,
            Sakiyama,S., Takahashi,H., Ohnuma,N., Tanabe,M., Fujimura,S. and
            Nakagawara,A.
    ITLE    GDNF-induced differentiation and its enhancement by retinoic acid
            in primary human neuroblastomas expressing c-Ret and GDNFR-alpha
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2175)
  AUTHORS   Nakagawara,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1997) Division of Biochemistry, Chiba Cancer
            Center Research Institute, 666-2, Nitona, Chuoh-ku, Chiba 260,
            Japan
FEATURES             Location/Qualifiers
     source          1. .2175
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /tissue_type="substantia nigra"
     CDS             245. .1627
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2459742"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISVEHIPKGNNCLDAAKACNLDD
                     ICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                     RRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKEN
                     YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTAL
                     STLLSLTETS"
BASE COUNT      544 a    566 c    535 g    516 t     14 others
ORIGIN
```

```
  Query Match           52.8%;  Score 1128;  DB 22;  Length 2175;
  Best Local Similarity 83.0%;  Pred. No. 0.00e+00;
  Matches  1634;  Conservative    0;  Mismatches 300;  Indels  35;  Gaps  17;

Db      1 TGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGC--CATCCCGGAGC 58
          ||||||||| |||||||||| |||||||  |||||||||||||||||  |  |||||| |
Qy     58 TGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCCGGATC 117

Db     59 TGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATGGA 117
          |||||||| |||||||||| | || |   |||  ||||| | ||||  | |||||||||
Qy    118 TGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGGATGGA 177

Db    118 GCTGAACTTTGGGCGGCCAGAGCAGCACAGCTGTCCGGGGATCGCTGCATGCTGAGCTCC 177
          |||||||||| | |||||||| ||| |||  | ||||||||||||||| |||||||||
Qy    178 GCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCTCT 237

Db    178 CTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTGGGGGGGCGGGGACCAGCCCCGCGC 237
          |||  | |||||  |||||||||  |||||||   ||||||||||||||||||  ||||
Qy    238 CTCCCCGAGACCGGGCGGCGGCTTTGGATTTT---GGGGGGGCGGGGACCAGCTGCGCGG 294

Db    238 CGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCTGTC 297
          |||||||||||||||| || || |||||||||||||||| ||| |||| |||||  ||||
Qy    295 CGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTC 354

Db    298 GGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCTGAA 357
          |||||  |||||  || |||||||| |||||  || ||||||||||  ||||||||||||
Qy    355 CGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAA 414

Db    358 GGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGAGAC 417
          ||| |||||||||||||||||||||||||| |||||||||||||||||||||||||||  ||
Qy    415 GGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAAC 474

Db    418 CAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGAGGC 477
          |||||||||||| |||||||||| |||||||||||||||||| ||||||||||||||||
Qy    475 CAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGC 534

Db    478 CCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAAGAA 537
          | |||||||||||||  || |||||||||||||||||||||||| ||||||||  ||||||||
Qy    535 CTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAA 594

Db    538 CTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGAGGA 597
          || ||||| || |||||||||||||||||||||||||||||||||||||| || ||||| ||
Qy    595 TTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGA 654

Db    598 TTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATTCAT 657
          |||||| |||| || |||||||||| |||||||||||||||||  ||||| |||||
Qy    655 TTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCAT 714

Db    658 ATCAG-TG-------G--AG-----CACATTCCCAAAGGGAACAACTGCCTGGATGCAGC 702
          |||||  ||       |  ||     |||||| ||||||||||||||||||||| |||||
Qy    715 ATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGC 774

Db    703 GAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTG 762
          |||||||||||||| |||||||  || ||||||||||||||||  |||||||||| ||
Qy    775 CAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTG 834

Db    763 CACCACCAGCGTGTCCAATGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCA 822
          |||||||||| ||||||| || |||||||||||||| ||||||||||||||||| ||||
Qy    835 CACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCA 894

Db    823 GTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGA 882
          |||||| |||||||| |||||||||||||||||| ||||||||||||||||||||||||
Qy    895 GTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGA 954

Db    883 CATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAG 942
          |||||||||||| |||||| |||||||||| ||||| || ||||||||||||||||  |
```

```
Qy    955 CATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACG 1014

Db    943 GGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAGATC 1002

Qy   1015 AGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATC 1074

Db   1003 TCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTGTCT 1062

Qy   1075 TCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCT 1134

Db   1063 AAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCATGAC 1122

Qy   1135 TAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGAC 1194

Db   1123 CCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAACAG 1182

Qy   1195 TCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAG 1254

Db   1183 TGGGAACGACCTAGAAGAGTGCTTGAAATTCTTGAATTTCTTCAAGGACAATACATGTCT 1242

Qy   1255 CGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCT 1314

Db   1243 TAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGCCTT 1302

Qy   1315 CAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCC 1374

Db   1303 CCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAACCCCT 1362

Qy   1375 TCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCT 1434

Db   1363 GGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAATTT 1422

Qy   1435 GGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTT 1494

Db   1423 ACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAATGG 1482

Qy   1495 GCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAG 1554

Db   1483 TAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAATGGC 1539

Qy   1555 TGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGC 1614

Db   1540 TGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTCCAC 1599

Qy   1615 TGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGC 1674

     1600 CCTATTATCTTTAACAGAAACATCATAGCTGCATTAAAAAAATACAATATGGACATGTAA 1659

     1675 CCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAA 1734

Db   1660 AAAGACAAA-AACCAAGTTATCTGTTTCCTGTTCTCTTGTATAGCTGAAATTCCAGTTT- 1717

Qy   1735 GACAAAAGAGAACCAAGTATTCTGTC-CCTGTCCTCTTGTATATCTGAAAATCCAGTTTT 1793

Db   1718 AGGAGCTCAGTTGAGAAACAGTTCCATTCAACTGGAACATTTTTTTTTTTTCCTTTTAAA 1777

Qy   1794 AAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAAC--TCTTTCCTTGT--TTTTAAG 1849

Db   1778 AAAGCTTCTTGTGATCCTTTGGGGCTTCTGTGAAAAACCTGATGCAGTGCTCCATCCAAA 1837

Qy   1850 AAAGCTTGT-G-GC-CCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAA 1906

Db   1838 CTCAAAAGGCTTNGGGATATGCTGTNTTTTAAAGGGACAGTTTGTAACTTGGGCTGTAAA 1897

Qy   1907 CCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAG 1966

Db   1898 NCAAACTGGGGNTGTGTTTTC-GATGATGATGATGATCATGATGATGAT 1945

Qy   1967 CTGGGCTTATCATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGAT 2015


RESULT   10
LOCUS       HSU97144     1707 bp    mRNA            PRI       29-JUL-1997
DEFINITION  Homo sapiens RET ligand 1 (RETL1) mRNA, complete cds.
ACCESSION   U97144
NID         g2282025
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 1707)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Glial cell line-derived neurotrophic factor-dependent RET
            activation can be mediated by two different cell-surface accessory
            proteins
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6238-6243 (1997)
  MEDLINE   97322356
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Sanicola,M., Hession,C.A., Worley,D.S., Carmillo,P., Ehrenfels,C.,
            Walus,L., Robinson,S., Jaworski,G., Wei,H., Tizard,R., Whitty,A.,
            Pepinsky,R.B. and Cate,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-1997) Molecular Genetics, BIOGEN, 14 Cambridge
            Center, Cambridge, MA 02142, USA
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="embryonic"
                     /tissue_type="kidney"
     gene            1. .1707
                     /gene="RETL1"
     CDS             262. .1644
                     /gene="RETL1"
                     /note="PI-linked cell-surface accessory protein; GDNFRa"
                     /codon_start=1
                     /product="RET ligand 1"
                     /db_xref="PID:g2282026"
                     /translation="MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCST
                     KYRTLRQCVAGKETNFSLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRI
                     YWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRVVPFISVEHIPKGNNCLDAAKACNLDD
                     ICKKYRSAYITPCTTSVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTE
                     RRRQTIVPVCSYEEREKPNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKEN
                     YADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLK
                     NAIQAFGNGSDVTVWQPAFPVQTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCAN
                     LQAQKLKSNVSGNTHLCISNGNYEKEGLGASSHITTKSMAAPPSCGLSPLLVLVVTAL
                     STLLSLTETS"
BASE COUNT      420 a    490 c    456 g    341 t
ORIGIN

  Query Match           51.1%;  Score 1093;  DB 22;  Length 1707;
  Best Local Similarity  87.2%;  Pred. No. 0.00e+00;
  Matches  1427;  Conservative    0;  Mismatches 184;  Indels  25;  Gaps   8;

Db      15 GTTGGGTCGGACCTGAACCCCTAAAAGCGGAACCGCCTCCCGCCCTCGC--CATCCCGGA 72

Qy      56 GTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGCCCGGA 115

Db      73 GCTGAGTCGCCGGCGGCGGTGG-CTGCTGCCAGACCCGGAGTTTCCTCTTTCACTGGATG 131

Qy     116 TCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCTGGATG 175

Db     132 GAGCTGAACTTTGGGCGGCCAGAACAGCACAGCTGTCCGGGGATCGCTGCACGCTGAGCT 191

Qy     176 GAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCTGAGCT 235

Db     192 CCCTCGGCAAGACCCAGCGGCGGCTCGGGATTTTTTTTGGGGGGGCGGGGACCAGCCCCG 251

Qy     236 CTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTT----GGGGGGGCGGGGACCAGCTGCG 291

Db     252 CGCCGGCACCATGTTCCTGGCGACCCTGTACTTCGCGCTGCCGCTCTTGGACTTGCTCCT 311
```

```
Qy    292 CGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGAT 351

Db    312 GTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGTGATCAGTGCCT 371

Qy    352 GTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCT 411

Db    372 GAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTGGCGGGCAAGGA 431

Qy    412 GAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGA 471

Db    432 GACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGCAGCGCCATGGA 491

Qy    472 AACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGA 531

Db    492 GGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATGAAGAAGGAGAA 551

Qy    532 GGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAA 591

Db    552 GAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGATCTGCTGGA 611

Qy    592 GAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGA 651

Db    612 GGATTCCCCATATGAACCAGTTAACAGCAGATTGTCAGATATATTCCGGGTGGTCCCATT 671

Qy    652 AGATTCCCCGTATGAGCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTT 711

Db    672 CATATCAG-TG-------G--AG-----CACATTCCCAAAGGGAACAACTGCCTGGATGC 716

Qy    712 CATATCAGATGTTTTCCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGC 771

Db    717 AGCGAAGGCCTGCAACCTCGACGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCC 776

Qy    772 AGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCC 831

Db    777 GTGCACCACCAGCGTGTCCAACGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCG 836

Qy    832 CTGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAG 891

Db    837 GCAGTTCTTTGACAAGGTCCCGGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCG 896

Qy    892 GCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 951

Db    897 GGACATCGCCTGCACAGAGCGGAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGA 956

Qy    952 GGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGA 1011

Db    957 GAGGGAGAAGCCCAACTGTTTGAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAG 1016

     1012 ACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAG 1071

     1017 ATCTCGCCTTGCGGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAGCTG 1076

Qy   1072 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 1131

Db   1077 TCTAAAGGAAAACTACGCTGACTGCCTCCTCGCCTACTCGGGGCTTATTGGCACAGTCAT 1136

Qy   1132 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 1191

Db   1137 GACCCCCAACTACATAGACTCCAGTAGCCTCAGTGTGGCCCCATGGTGTGACTGCAGCAA 1196

Qy   1192 GACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAA 1251

Db   1197 CAGTGGGAACGACCTAGAAGAGTGCTTGAAATTTTTGAATTTCTTCAAGGACAATACATG 1256

Qy   1252 CAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTG 1311

Db   1257 TCTTAAAAATGCAATTCAAGCCTTTGGCAATGGCTCCGATGTGACCGTGTGGCAGCCAGC 1316

Qy   1312 TCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGC 1371

Db   1317 CTTCCCAGTACAGACCACCACTGCCACTACCACCACTGCCCTCCGGGTTAAGAACAAGCC 1376

Qy   1372 CCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCC 1431

Db   1377 CCTGGGGCCAGCAGGGTCTGAGAATGAAATTCCCACTCATGTTTTGCCACCGTGTGCAAA 1436

Qy   1432 TCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAA 1491

Db   1437 TTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAATACACACCTCTGTATTTCCAA 1496

Qy   1492 TTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGA 1551

Db   1497 TGGTAATTATGAAAAAGAAGGTCTCG---GTGCTTCCAGCCACATAACCACAAAATCAAT 1553

Qy   1552 TAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAAT 1611

Db   1554 GGCTGCTCCTCCAAGCTGTGGTCTGAGCCCACTGCTGGTCCTGGTGGTAACCGCTCTGTC 1613

Qy   1612 GGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGC 1671

Db   1614 CACCCTATTATCTTTA 1629

Qy   1672 TGCCCTGTTATCTGTA 1687
```

```
RESULT   11
LOCUS       AF012811      777 bp    mRNA            ROD       06-AUG-1997
DEFINITION  Mus musculus GDNF receptor alpha mRNA, partial cds.
ACCESSION   AF012811
NID         g2286224
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   Puliti,A. and Seri,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Laboratorio di Genetica Molecolare,
            Istituto Giannina Gaslini, Largo G. Gaslini, 5, Genova 16148, Italy
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     source          1. .430
                     /organism="Mus musculus"
                     /clone="402011"
                     /note="EST deposited under GenBank Accession Number W76716
                     from Soares mouse embryo NbME13.5 14.5 cDNA library; IMAGE
                     Consortium"
     CDS             <1. .680
                     /codon_start=3
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2286225"
                     /translation="LNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLL
                     AYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAF
                     GNGSDVTMWQPAPPVQTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKL
                     KSNVSGSTHLCLSDNDYGKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLS
                     VSLAETS"
BASE COUNT      202 a    212 c    174 g    189 t
ORIGIN
```

```
  Query Match           29.9%;  Score 640;  DB 23;  Length 777;
  Best Local Similarity  93.6%;  Pred. No. 0.00e+00;
  Matches   735;  Conservative    0;  Mismatches  41;  Indels   9;  Gaps   2;

Db      1 GCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATT 60

Qy   1029 GCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATT 1088

Db     61 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 120

Qy   1089 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 1148

Db    121 CAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAG 180
```

```
Qy   1149 CAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAG 1208

Db    181 ACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGG 240
Qy   1209 ACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGG 1268

Db    241 AAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTC 300
Qy   1269 AAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTC 1328

Db    301 AAGCCTTTGGCAATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCCAGTCCAGACCA 360
Qy   1329 AAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCA 1388

Db    361 CCACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCT 420
Qy   1389 CCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGT 1448

Db    421 CTGAGAATGAGATTCCCACACACGTTTTACCACCGTGTGCTAATTTGCAGGCACAGAAGC 480
Qy   1449 CTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAAGC 1508

Db    481 TGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTTTCTGATAATGATTACGGAAAGG 540
Qy   1509 TGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGGAAAGG 1568

Db    541 ATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCT 600
Qy   1569 ATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCT 1628

Db    601 GCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCTCTGGCTGCCCTGTTGTCTGTAT 660
Qy   1629 GCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTATCTGTAT 1688

Db    661 CATTGGCAGAAACATCGTAGCTGCATCC--------AGTATGAAAAGACAAAAGAAAACC 712
Qy   1689 CGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGACAAAAGAGAACC 1748

Db    713 AAGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCTCGGTTGA 772
Qy   1749 AAGTATTCTGTCCC-TGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCTCCGTTGA 1807

Db    773 GAAGC 777
Qy   1808 GAAGC 1812
```

[redacted]LT 12

```
LOCUS       GGU90541     3037 bp    mRNA            VRT       25-JUN-1997
DEFINITION  Gallus gallus GDNF receptor alpha (GDNFRalpha) mRNA, complete cds.
ACCESSION   U90541
NID         g2213802
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
            Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 3037)
  AUTHORS   Buj-Bello,A., Adu,J., Pinon,L.G., Horton,A., Thompson,J.,
            Rosenthal,A., Chinchetru,M., Buchman,V.L. and Davies,A.M.
  TITLE     Neurturin responsiveness requires a GPI-linked receptor and the Ret
            receptor tyrosine kinase
  JOURNAL   Nature 387 (6634), 721-724 (1997)
  MEDLINE   97336104
REFERENCE   2  (bases 1 to 3037)
  AUTHORS   Buchman,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1997) Biol. Med. Sciences, Univ. of St. Andrews,
            Bute Medical Buildings, St. Andrews, Fife KY16 9TS, Scotland
FEATURES             Location/Qualifiers
     source          1. .3037
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            740. .2150
                     /gene="GDNFRalpha"
     CDS             740. .2149
                     /gene="GDNFRalpha"
                     /note="GPI-linked component of GDNF receptor"
                     /codon_start=1
                     /product="GDNF receptor alpha"
                     /db_xref="PID:g2213803"
                     /translation="MFLALLYLALPLADVLLSAEVSGLPGGDRLDCVKASDQCLKEQS
                     CSTKYRTLRQCVAGKESNFSRATGLEAKDECKSAMEALKQKSLYNCRCKRGMKKEKNC
                     LRIYWSMYQSLQGNDLLEDSPYEPVNSRLSDIFRLAPIVSVEPVLSKGNNCLDAAKAC
                     NLNDTCKRFRSAYITPCTSSTSNEICNKRKCHKALRLFFDKVPPKHSYGMLFCSCRDV
                     ACTERRRQTIVPVCSYEDREKPNCLNLQESCKKNYICRSRLADFFTNCQPESRSVSSC
                     LKENYADCLLAYSGLIGTVMTPNYIDSSSLSVAPWCDCSNSGNDIDECRKFLNFFQDN
                     TCLKNAIQAFGNGTDVNVWQPILPVQTTTATTTTASRLKNTGSETTNNEIPTHNDSPA
                     CANLQAQKKRKSNESVDTELCLNENAIGKDNTPGVSTSHISSENSFALPTSFYPSTPL
                     ILMTIALSLFLFLSSSVVL"
     polyA_signal    3005. .3010
BASE COUNT      773 a    764 c    722 g    778 t
ORIGIN
```

Query Match 27.0%; Score 578; DB 17; Length 3037;
Best Local Similarity 78.1%; Pred. No. 0.00e+00;
Matches 942; Conservative 0; Mismatches 244; Indels 20; Gaps 10;

```
Db    815 GGAGGGGACCGCCTCGACTGCGTGAAAGCCAGCGATCAGTGTCTCAAGGAGCAGAGCTGT 874
Qy    368 GGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAGAGCTGC 427

Db    875 AGTACTAAGTACAGGACACTGAGGCAGTGTGTAGCCGGCAAAGAGAGCAACTTCAGCCGG 934
Qy    428 AGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTG 487

Db    935 GCGACGGGCCTGGAGGCGAAGGATGAATGCAAAAGCGCCATGGAGGCTCTCAAGCAGAAA 994
Qy    488 ACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAGCAGAAG 547

Db    995 TCTCTGTACAACTGCCGCTGCAAGAGGGGCATGAAGAAGGAGAAAAACTGCCTGCGCATT 1054
Qy    548 TCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATC 607

Db   1055 TACTGGAGCATGTACCAGAGCTTACAGGGAAATGACTTGCTTGAGGATTCTCCCTATGAA 1114
Qy    608 TACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCGTATGAG 667

Db   1115 CCAGTTAACAGCAGGCTATCAGACATATTCAGGCTAGCACCGATTGTATCAG-TGG---- 1169
Qy    668 CCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTTC 727

Db   1170 -AGCCAGT-----AC-TATCAAAGGGGAACAACTGCTTGGATGCAGCAAAAGCTTGTAAC 1222
Qy    728 CAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAAC 787

Db   1223 CTAAATGATACCTGCAAGAGGTTCAGATCTGCTTACATAACCCCCTGCACCAGCAGCACG 1282
Qy    788 CTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACCAGCATG 847

Db   1283 TCTAATGAAATCTGTAACAAGCGGAAGTGTCATAAGGCCCTCCGGCTATTTTTTGACAAA 1342
Qy    848 TCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAG 907

Db   1343 GTTCCCCCAAAGCACAGCTACGGGATGCTCTTCTGCTCCTGTCGAGACGTAGCCTGTACA 1402
Qy    908 GTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACC 967

Db   1403 GAAAGGAGGCGGCAGACTATTGTTCCTGTGTGTTCATATGAGGACAGGGAGAAACCAAAC 1462
Qy    968 GAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAAC 1027

Db   1463 TGCCTGAATTTACAAGAATCCTGCAAGAAGAATTACATCTGCAGATCTCGCCTTGCAGAT 1522
```

```
Qy   1028 TGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGAT 1087

Db   1523 TTTTTCACAAACTGCCAGCCTGAGTCACGCTCTGTTAGTAGCTGTCTGAAGGAGAACTAC 1582
          ||||| || |||||||||||| |||||| | ||||| || | |||||| ||||||||||||
Qy   1088 TTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTAC 1147

Db   1583 GCTGACTGCCTCCTCGCTTACTCGGGGCTCATTGGCACAGTGATGACACCGAACTACATA 1642
          || |||||||||| || |||||||| || |||||||||| ||||| || |||||| ||
Qy   1148 GCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTA 1207

Db   1643 GACTCAAGTAGTCTCAGCGTTGCCCCATGGTGTGACTGCAGCAACAGTGGTAATGACATA 1702
          ||||| || || |||||||| || ||||||||||||||||||||||| || |||||| |
Qy   1208 GACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTG 1267

Db   1703 GATGAATGCCGGAAATTTCTGAATTTCTTCCAGGACAACACATGCCTTAAAAATGCAATT 1762
          || || |||  |||||||||||||| ||  ||||||| || || || ||||||||||||
Qy   1268 GAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATT 1327

Db   1763 CAGGCTTTCGGCAATGGTACTGATGTGAATGTGTGGCAGCCAATATTACCGGTACAGACC 1822
          || || || |||||||| | |||||||   |||||||||||      || || ||||||
Qy   1328 CAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACC 1387

Db   1823 ACTACAGCCACAACTACAACAGCTTCCAGACTTAAAAACACAGGTTCAGAGACCACCAAC 1882
          || || ||||| ||||| || || | | |  | || |||| ||  || | | ||| ||
Qy   1388 ACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACA-AGCCTCTGGGGCCAGCAGG 1446

Db   1883 AA-TGA-AAT-ACCCACCCACA-ATGATTCACCAGCATGTGCGAACCTGCAGGCACAGAA 1938
            ||| ||| |    |||||| | | || |||| | ||||||||  ||||||| |||||
Qy   1447 GTCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAA 1506

Db   1939 GAAGCGGAAATCCAATGAATCTGTAGATACAGAACTCTGTCTTAATGAGAATGCTATTGG 1998
          |   | ||||||||||  || |    ||| | |||||||||  ||| | || | | ||
Qy   1507 G---CTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGG 1563

Db   1999 GAAAGA 2004
            || ||
Qy   1564 AAAGGA 1569
```

```
RESULT   13
LOCUS       HSGFRA1G06    596 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) gene, exon 6.
ACCESSION   AF038415
NID         g2921538
KEYWORDS    .
SEGMENT     6 of 11
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 596)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
     exon            147. .483
                     /gene="GFRA1"
                     /number=6
BASE COUNT      152 a    156 c    163 g    125 t
ORIGIN
```

```
  Query Match           12.9%;  Score 275;  DB 22;  Length 596;
  Best Local Similarity  90.6%;  Pred. No. 1.19e-205;
  Matches   307;  Conservative    0;  Mismatches  32;  Indels   0;  Gaps   0;

Db    145 AGTGGAGCACATTCCCAAAGGGAACAACTGCCTGGATGCAGCGAAGGCCTGCAACCTCGA 204
          |||||| |||||| |||||||||||||||||||||| ||||| |||||||||||||| ||
Qy    733 AGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGA 792

Db    205 CGACATTTGCAAGAAGTACAGGTCGGCGTACATCACCCCGTGCACCACCAGCGTGTCCAA 264
          |||||  || |||||||||||||||| |||||||||| |||||||||||| |||||||
Qy    793 CGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACCAGCATGTCCAA 852

Db    265 CGATGTCTGCAACCGCCGCAAGTGCCACAAGGCCCTCCGGCAGTTCTTTGACAAGGTCCC 324
          ||| |||||||||||||| |||||||||||||||||| ||||||||| |||||||| ||
Qy    853 CGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCC 912

Db    325 GGCCAAGCACAGCTACGGAATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACAGAGCG 384
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||| |||||
Qy    913 GGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCG 972

Db    385 GAGGCGACAGACCATCGTGCCTGTGTGCTCCTATGAAGAGAGGGAGAAGCCCAACTGTTT 444
          | |||||||||| ||||| || |||||||||||||||||  | |||| |||||||||  |
Qy    973 GCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCT 1032

Db    445 GAATTTGCAGGACTCCTGCAAGACGAATTACATCTGCAG 483
          || | |||| |||||||||||||| ||||||||||||||
Qy   1033 GAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAG 1071
```

```
RESULT   14
LOCUS       HSGFRA1G03    497 bp    DNA             PRI       10-APR-1998
DEFINITION  Homo sapiens GPI-linked anchor protein (GFRA1) gene, exon 3.
ACCESSION   AF038412
NID         g2921535
KEYWORDS    .
SEGMENT     3 of 11
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Human GFRA1: Cloning, mapping, genomic structure, and evaluation as
            a candidate gene for Hirschsprung disease susceptibility
  JOURNAL   Genomics 48, 354-362 (1998)
REFERENCE   2  (bases 1 to 497)
  AUTHORS   Angrist,M., Jing,S., Bolk,S., Bentley,K., Nallasamy,S.,
            Halushka,M., Fox,G.M. and Chakravarti,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genetics, Case Western Reserve, 2109
            Adelbert Road, Cleveland, OH 44106-4955, USA
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10q25"
     exon            76. .369
                     /gene="GFRA1"
                     /number=3
BASE COUNT       94 a    174 c    151 g     78 t
ORIGIN
```

```
  Query Match           11.4%;  Score 244;  DB 22;  Length 497;
  Best Local Similarity  91.2%;  Pred. No. 2.41e-178;
  Matches   270;  Conservative    0;  Mismatches  26;  Indels   0;  Gaps   0;

Db     75 GACTTGCTCCTGTCGGCCGAAGTGAGCGGCGGAGACCGCCTGGATTGCGTGAAAGCCAGT 134
          || |||||  |||| ||||| ||||| || |||||||| ||||| || ||||||||||
Qy    341 GATTTGCTGATGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGC 400
```

```
Db    135 GATCAGTGCCTGAAGGAGCAGAGCTGCAGCACCAAGTACCGCACGCTAAGGCAGTGCGTG 194
Qy    401 GATCAGTGCCTGAAGGAACAGAGCTGCAGCACCAAGTACCGCACACTAAGGCAGTGCGTG 460

Db    195 GCGGGCAAGGAGACCAACTTCAGCCTGGCATCCGGCCTGGAGGCCAAGGATGAGTGCCGC 254
Qy    461 GCGGGCAAGGAAACCAACTTCAGCCTGACATCCGGCCTTGAGGCCAAGGATGAGTGCCGT 520

Db    255 AGCGCCATGGAGGCCCTGAAGCAGAAGTCGCTCTACAACTGCCGCTGCAAGCGGGGTATG 314
Qy    521 AGCGCCATGGAGGCCTTGAAGCAGAAGTCTCTGTACAACTGCCGCTGCAAGCGGGGCATG 580

Db    315 AAGAAGGAGAAGAACTGCCTGCGCATTTACTGGAGCATGTACCAGAGCCTGCAGGG 370
Qy    581 AAGAAAGAGAAGAATTGTCTGCGTATCTACTGGAGCATGTACCAGAGCCTGCAGGG 636


RESULT   15
LOCUS       HSU93703     1395 bp    mRNA            PRI       01-JUL-1997
DEFINITION  Human glial cell line-derived neurotrophic factor receptor beta
            (GDNFR-beta) mRNA, complete cds.
ACCESSION   U93703
NID         g2228736
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 1395)
  AUTHORS   Wartiovaara K., Suvanto P., Horelli-Kuitunen N., Lindahl M.,
            Moshnyakov M., Airaksinen M.S., Palotie A., Sariola H. and Saarma
            M.
  TITLE     Cloning, mRNA distribution and chromosomal localisation of the gene
            for the human glial cell line-derived neurotrophic factor beta, a
            homologue to GDNFR-alpha
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1395)
  AUTHORS   Suvanto P., Wartiovaara K. and Saarma M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1997) Molecular Neurobiology, Institute of
            Biotechnology, JP 56, University of Helsinki 00014, Finland
FEATURES             Location/Qualifiers
     source          1. .1395
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p21-p22"
                     /tissue_type="fetal brain"
     gene            1. .1395
                     /gene="GDNFR-beta"
     primer_bind     1. .23
                     /gene="GDNFR-beta"
                     /note="forward PCR primer; designed from rat GDNFR-beta
                     mRNA"
     CDS             1. .1395
                     /gene="GDNFR-beta"
                     /codon_start=1
                     /product="glial cell line-derived neurotrophic factor
                     receptor beta"
                     /db_xref="PID:g2228737"
                     /translation="MILANAFCLFFFLDETLRSLASPSSLQGPELHGWRPPVDCVRAN
                     ELCAAESNCSSRYRTLRQCLAGRDRNTMLANKECQAALEVLQESPLYDCRCKRGMKKE
                     LQCLQIYWSIHLGLTEGEEFYEASPYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNH
                     CLDAAKACNLNDNCKKLRSSYISICNREISPTERCNRRKCHKALRQFFDRVPSEYTYR
                     MLFCSCQDQACAERRRQTILPSCSYEDKEKPNCLDLRGVCRTDHLCRSRLADFHANCR
                     ASYQTVTSCPADNYQACLGSYAGMIGFDMTPNYVDSSPTGIVVSPWCSCRGSGNMEEE
                     CEKFLRDFTENPCLRNAIQAFGNGTDVNVSPKGPSFQATQAPRVEKTPSLPDDLSDST
                     SLGTSVITTCTSVQEQGLKANNSKELSMCFTELTTNIIPGSNKVIKPNSGPSRARPSA
                     ALTVLSVLMLKQAL"
BASE COUNT      281 a    463 c    389 g    262 t
ORIGIN
```

```
  Query Match          10.5%;  Score 225;  DB 22;  Length 1395;
  Best Local Similarity  65.8%;  Pred. No. 1.09e-161;
  Matches   593;  Conservative    0;  Mismatches 296;  Indels  12;  Gaps   8;

Db    219 GGCCAACAAGGAGTGCCAGGCGGCCTTGGAGGTCTTGCAGGAGAGCCCGCTGTACGACTG 278
Qy    502 GGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAGCAGAAGTCTCTGTACAACTG 561

Db    279 CCGCTGCAAGCGGGGCATGAAGAAGGAGCTGCAGTGTCTGCAGATCTACTGGAGCATCCA 338
Qy    562 CCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATCTACTGGAGCATGTA 621

Db    339 CCTGGGGCTGACCGAGGGTGAGGAGTTCTACGAAGCCTCCCCCTATGAGCCGGTGACCTC 398
Qy    622 CCAGAGCCTG--C-AGGGAAATGACCTCCTGGAAGATTCCCCGTATGAGCCGGTTAACAG 678

Db    399 CCGCCTCTCGGACATCTTCAGGCTTGCTTCAATCTTCTCAGGGACAGGGGCAGACCCGGT 458
Qy    679 CAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTTCCAGCAAGTGGA 738

Db    459 GGTCAGCGCCAAGAGCAACCATTGCCTGGATGCTGCCAAGGCCTGCAACCTGAATGACAA 518
Qy    739 ACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACAC 798

Db    519 CTGCAAGAAGCTGCGCTCCTCCTACATCTCCATCTGCAACCGCGAGATCTCGCCCACCGA 578
Qy    799 CTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGA 855

Db    579 GCGCTGCAACCGCCGCAAGTGCCACAAGGCCCTGCGCCAGTTCTTCGACCGGGTGCCCAG 638
Qy    856 GGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGC 915

Db    639 CGAGTACACCTACCGCATGCTCTTCTGCTCCTGCCAAGACCAGGCGTGCGCTGAGCGCCG 698
Qy    916 CAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCG 975

Db    699 CCGGCAAACCATCCTGCCCAGCTGCTCCTATGAGGACAAGGAGAAGCCCAACTGCCTGGA 758
Qy    976 GCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAG 1035

Db    759 CCTGCGTGGCGTGTGCCGGACTGACCACCTGTGTCGGTCCCGGCTGGCCGACTTCCATGC 818
Qy   1036 TCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATTTTTTTAC 1095

Db    819 CAATTGTCGAGCCTCCTACCAGACGGTCACCAGCTGCCCTGCGGACAATTACCAGGCGTG 878
Qy   1096 CAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGACTG 1155

Db    879 TCTGGGCTCTTATGCTGGCATGATTGGGTTTGACATGACACCTAACTATGTGGACTCCAG 938
Qy   1156 CCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAGACTCCAG 1215

Db    939 CCCCACTGGCATCGTGGTGTCCCCCTGGTGCAGCTGTCGTGGCAGCGGGAACATGGAGGA 998
Qy   1216 CAGC-CT--CAGCGTGG---CACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGGA 1269

Db    999 GGAGTGTGAGAAGTTCCTCAGGGACTTCACCGAGAACCCATGCCTCCGGAACGCCATCCA 1058
Qy   1270 AGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTCA 1329

Db   1059 GGCCTTTGGCAACGGCACGGACGTGAACGTGTCCCCAAAAGGCCCCTCGTTCCAGGCCAC 1118
Qy   1330 AGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCAC 1389

Db   1119 C 1119
Qy   1390 C 1390


Search completed: Wed Jun 24 22:01:39 1998
Job time : 4788 secs.
```

****

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

[redacted] on: Wed Jun 24 23:10:17 1998; MasPar time 314.60 Seconds
865.544 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-866-354-3
Description: (1-2138) from US08866354.seq
Perfect Score: 2138
N.A. Sequence: 1 AGCTCGCTCTCCCGGGGCAG.....AGACAGGCCCGCAGCCGTCG 2138
Comp: TCGAGCGAGAGGGCCCCGTC.....TCTGTCCGGGCGTCGGCAGC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 176886 seqs, 63680241 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37

tistics: Mean 9.707; Variance 5.482; scale 1.771

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	1885	88.2	2378	37	T84975	Rat glial cell derive	0.00e+00
	2	748	35.0	840	37	T84977	Mouse GDNFR alpha clo	0.00e+00
	3	347	16.2	418	37	T84976	Mouse GDNFR alpha clo	2.00e-230
	4	133	6.2	453	37	T84979	Human EST-derived seq	1.41e-72
	5	122	5.7	351	37	T84978	Human EST-derived seq	1.01e-64
	6	112	5.2	385	15	Q85526	DNA probe 32 detects	1.25e-57
	7	96	4.5	201	37	T84981	Human EST-derived seq	2.10e-46
	8	96	4.5	201	37	T84980	Human EST-derived seq	2.10e-46
	9	44	2.1	91	9	Q51746	Oligonucleotide probe	7.63e-12
c	10	44	2.1	91	9	Q51746	Oligonucleotide probe	7.63e-12
c	11	43	2.0	204	1	N81164	Base substituted E.co	3.11e-11
	12	40	1.9	204	1	N81164	Base substituted E.co	1.99e-09
	13	38	1.8	114	12	Q70468	Generic DNA sequence	3.05e-08
	14	38	1.8	114	12	Q70465	Generic DNA sequence	3.05e-08
	15	38	1.8	114	12	Q70469	Generic DNA sequence	3.05e-08
	16	39	1.8	172	32	T76363	Human interleukin 8 a	7.83e-09
	17	36	1.7	114	12	Q70467	Generic DNA sequence	4.46e-07
	18	36	1.7	114	12	Q70466	Generic DNA sequence	4.46e-07
	19	36	1.7	190	31	T76452	Chymase antisense oli	4.46e-07
	20	34	1.6	114	12	Q70470	Generic DNA sequence	6.22e-06
	21	34	1.6	114	12	Q70473	Generic DNA sequence	6.22e-06
c	22	35	1.6	114	12	Q70467	Generic DNA sequence	1.67e-06
c	23	34	1.6	114	12	Q70468	Generic DNA sequence	6.22e-06
c	24	34	1.6	114	12	Q70469	Generic DNA sequence	6.22e-06
c	25	34	1.6	114	12	Q70465	Generic DNA sequence	6.22e-06
	26	33	1.5	114	12	Q70471	Generic DNA sequence	2.28e-05
	27	33	1.5	114	12	Q70472	Generic DNA sequence	2.28e-05
c	28	33	1.5	114	12	Q70470	Generic DNA sequence	2.28e-05
c	29	32	1.5	114	12	Q70466	Generic DNA sequence	8.22e-05
c	30	32	1.5	168	31	T76270	Human MDNCF antisense	8.22e-05
c	31	33	1.5	172	32	T76363	Human interleukin 8 a	2.28e-05
	32	32	1.5	178	31	T76405	Human endothelin-1 an	8.22e-05
c	33	29	1.4	42	3	Q14326	MCPC 603 VH CDR2 walk	3.52e-03
	34	30	1.4	100	31	T76186	Human IL4 receptor an	1.02e-03
c	35	29	1.4	114	12	Q70472	Generic DNA sequence	3.52e-03
c	36	30	1.4	178	31	T76405	Human endothelin-1 an	1.02e-03
c	37	31	1.4	250	31	T76438	Substance P antisense	2.92e-04
	38	31	1.4	3871	2	N71302	HSV-1 gB and surround	2.92e-04
	39	27	1.3	39	7	Q51787	Mixed oligonucleotide	3.95e-02
c	40	27	1.3	114	12	Q70473	Generic DNA sequence	3.95e-02
c	41	27	1.3	128	31	T76233	Human IL6 antisense o	3.95e-02
	42	28	1.3	162	31	T76307	Human RANTES antisens	1.19e-02
c	43	28	1.3	190	31	T76452	Chymase antisense oli	1.19e-02
	44	28	1.3	264	31	T76445	Substance P receptor	1.19e-02
c	45	27	1.3	264	31	T76445	Substance P receptor	3.95e-02

ALIGNMENTS

RESULT 1
ID T84975 standard; cDNA; 2378 BP.
AC T84975;
DT 27-APR-1998 (first entry)
DE Rat glial cell derived neurotrophic factor receptor alpha cDNA.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; rat; kidney disease; glomerulonephritis; therapy; ds.
OS Rattus sp.
FH Key Location/Qualifiers
FT CDS 117..1523
FT /*tag= a
FT sig_peptide 117..188
FT /*tag= b
FT mat_peptide 189..1520
FT /*tag= c
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
DR P-PSDB; W27327.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Claim 23; Page 74-77; 100pp; English.
CC This cDNA codes for full-length rat glial cell derived neurotrophic
CC factor receptor alpha (GDNFR alpha) (see W27327), a novel
CC GPI-linked protein that is a ligand-binding component of the
CC receptor system for GDNF. It was isolated by expression cloning.
CC A cDNA library obtained from ventral midbrain tissue of E14 rat
CC embryos was generated in a cytomegalovirus-based vector. cDNA
CC clones were transfected into COS 7 cells and expression of
CC putative GDNF receptors was detected by binding of iodinated GDNF.
CC A single positive pool was obtained, from which the cDNA clone was
CC isolated. An expression vector containing the cDNA can be used to

```
CC   produce GDNFR alpha in transformed host cells.  The invention
CC   relates to novel uses of GDNF and its receptor.  In particular, it
CC   relates to native rat GDNFR alpha (see W27327), its variants and
CC   soluble derivatives (extracellular domain), chimeric GDNFR alpha
CC   and antibodies which bind to the GDNFR alpha, including agonist
CC   and neutralising antibodies, as well as various uses for these
CC   molecules.  It also relates to assay systems for detecting ligands
CC   to GDNFR alpha, systems for studying the physiological role of
CC   GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC   methods for identifying molecules homologous to GDNFR alpha, and
CC   therapeutic techniques (claimed) for the treatment of GDNF-related
CC   and GDNFR alpha-related conditions, particularly kidney disease
CC   associated with glomerulonephritis and enteric nervous system
CC   related disorders.  Transgenic and knockout animals are also
CC   claimed.
SQ   Sequence   2378 BP;    617 A;    624 C;    598 G;    539 T;

  Query Match           88.2%;  Score 1885;  DB 37;  Length 2378;
  Best Local Similarity  99.9%;  Pred. No. 0.00e+00;
  Matches  1886;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db      63 ccgggcggcggctttggattttgggggggcggggaccagctgcgcggcggcaccatgttc 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     248 CCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGCGCGGCGGCACCATGTTC 307

Db     123 ctagccactctgtacttcgcgctgccactcctggatttgctgatgtccgccgaggtgagt 182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     308 CTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGATGTCCGCCGAGGTGAGT 367

Db     183 ggtggagaccgtctggactgtgtgaaagccagcgatcagtgcctgaaggaacagagctgc 242
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     368 GGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCCTGAAGGAACAGAGCTGC 427

Db     243 agcaccaagtaccgcacactaaggcagtgcgtggcgggcaaggaaaccaacttcagcctg 302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     428 AGCACCAAGTACCGCACACTAAGGCAGTGCGTGGCGGGCAAGGAAACCAACTTCAGCCTG 487

Db     303 acatccggccttgaggccaaggatgagtgccgtagcgccatggaggccttgaagcagaag 362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     488 ACATCCGGCCTTGAGGCCAAGGATGAGTGCCGTAGCGCCATGGAGGCCTTGAAGCAGAAG 547

Db     363 tctctgtacaactgccgctgcaagcggggcatgaagaaagagaagaattgtctgcgtatc 422
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     548 TCTCTGTACAACTGCCGCTGCAAGCGGGGCATGAAGAAAGAGAAGAATTGTCTGCGTATC 607

Db     423 tactggagcatgtaccagagcctgcagggaaatgacctcctggaagattccccgtatgag 482
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       608 TACTGGAGCATGTACCAGAGCCTGCAGGGAAATGACCTCCTGGAAGATTCCCCGTATGAG 667

       483 ccggttaacagcaggttgtcagatatattccgggcagtcccgttcatatcagatgttttc 542
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     668 CCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTTC 727

Db     543 cagcaagtggaacacatttccaaagggaacaactgcctggacgcagccaaggcctgcaac 602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     728 CAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAAC 787

Db     603 ctggacgacacctgtaagaagtacaggtcggcctacatcacccccctgcaccaccagcatg 662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     788 CTGGACGACACCTGTAAGAAGTACAGGTCGGCCTACATCACCCCCTGCACCACCAGCATG 847

Db     663 tccaacgaggtctgcaaccgccgtaagtgccacaaggccctcaggcagttcttcgacaag 722
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     848 TCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAG 907

Db     723 gttccggccaagcacagctacgggatgctcttctgctcctgccgggacatcgcctgcacc 782
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     908 GTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACC 967

Db     783 gagcggcggcgacagactatcgtccccgtgtgctcctatgaagaacgagagaggcccaac 842
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     968 GAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAAC 1027

Db     843 tgcctgagtctgcaagactcctgcaagaccaattatatctgcagatctcgccttgcagat 902
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Qy    1028 TGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGAT 1087

Db     903 ttttttaccaactgccagccagagtcaaggtctgtcagcaactgtcttaaggagaactac 962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1088 TTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTAC 1147

Db     963 gcagactgcctcctggcctactcgggactgattggcacagtcatgactcccaactacgta 1022
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1148 GCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTA 1207

Db    1023 gactccagcagcctcagcgtggcaccatggtgtgactgcagcaacagcggcaatgacctg 1082
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1208 GACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTG 1267

Db    1083 gaagactgcttgaaatttctgaatttttttaaggacaatacttgtctcaaaaatgcaatt 1142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1268 GAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATT 1327

Db    1143 caagcctttggcaatggctcagatgtgaccatgtggcagccagcccctccagtccagacc 1202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1328 CAAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACC 1387

Db    1203 accactgccaccactaccactgccttccgggtcaagaacaagcctctggggccagcaggg 1262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1388 ACCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGG 1447

Db    1263 tctgagaatgagatccccacacacgttttaccaccctgtgcgaatttgcaggctcagaag 1322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1448 TCTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAAG 1507

Db    1323 ctgaaatccaatgtgtcgggtagcacacacctctgtctttctgatagtgatttcggaaag 1382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1508 CTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGGAAAG 1567

Db    1383 gatggtctcgctggtgcctccagccacataaccacaaaatcaatggctgctcctcccagc 1442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1568 GATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGC 1627

Db    1443 tgcagtctgagctcactgccggtgctgatgctcaccgcccttgctgccctgttatctgta 1502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1628 TGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTATCTGTA 1687

Db    1503 tcgttggcagaaacgtcgtagctgcatccgggaaaacagtatgaaaagacaaaagagaac 1562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1688 TCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGACAAAAGAGAAC 1747

Db    1563 caagtattctgtccctgtcctcttgtatatctgaaaatccagttttaaaagctccgttga 1622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1748 CAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATCCAGTTTTAAAAGCTCCGTTGA 1807

Db    1623 gaagcagtttcacccaactggaactctttccttgtttttaagaaagcttgtggccctcag 1682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1808 GAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAAAGCTTGTGGCCCTCAG 1867

Db    1683 gggcttctgttgaagaactgctacagggctaattccaaacccataaggctctggggcgtg 1742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1868 GGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCATAAGGCTCTGGGGCGTG 1927

Db    1743 gtgcggcttaaggggaccatttgcaccatgtaaagcaagctgggcttatcatgtgtttga 1802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1928 GTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGGCTTATCATGTGTTTGA 1987

Db    1803 tggtgaggatggtagtggtgatgatgatggtaattttaacagcttgaaccctgttctctc 1862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1988 TGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCTTGAACCCTGTTCTCTC 2047

Db    1863 tactggttaggaacaggagatactattgataaagattcttccatgtcttactcagcagca 1922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2048 TACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCATGTCTTACTCAGCAGCA 2107

Db    1923 ttgccttctgaagacaggcccgcagcc 1949
```

```
          ||||||||||||||||||||||||||
Qy   2108 TTGCCTTCTGAAGACAGGCCCGCAGCC 2134


RESULT    2
ID    T84977 standard; cDNA; 840 BP.
AC    T84977;
DT    27-APR-1998  (first entry)
DE    Mouse GDNFR alpha clone 26 3' end.
KW    Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW    GDNF; mouse; kidney disease; glomerulonephritis; therapy; ss.
OS    Mus musculus.
PN    WO9733912-A2.
PD    18-SEP-1997.
PF    13-MAR-1997; U04363.
PR    14-MAR-1996; US-618236.
PR    14-MAR-1996; US-615902.
PA    (GETH ) GENENTECH INC.
PI    Klein RD, Moore MW, Rosenthal A, Ryan AM;
      WPI; 97-470819/43.
      Isolated glial cell derived neurotrophic factor receptor alpha -
      useful to develop products to diagnose and treat associated
PT    disorders, particularly enteric nervous system or kidney disorders
PS    Example 1; Page 60-61; 100pp; English.
CC    This cDNA sequence comprises the 3' end of mouse full-length
CC    glial cell derived neurotrophic factor receptor alpha (GDNFR alpha)
CC    (see W27327) clone 26; the 5' end of the clone is given in
CC    T84978.  Clone 26 was isolated from a mouse cDNA library using
CC    rat GDNFR alpha cDNA (see T84975) as probe.  The invention relates
CC    to novel uses of GDNF and its receptor.  In particular, it relates
CC    to native rat GDNFR alpha (see W27327), its variants and soluble
CC    derivatives (extracellular domain), chimeric GDNFR alpha and
CC    antibodies which bind to the GDNFR alpha, including agonist and
CC    neutralising antibodies, as well as various uses for these
CC    molecules.  It also relates to assay systems for detecting ligands
CC    to GDNFR alpha, systems for studying the physiological role of
CC    GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC    methods for identifying molecules homologous to GDNFR alpha, and
CC    therapeutic techniques (claimed) for the treatment of GDNF-related
CC    and GDNFR alpha-related conditions, particularly kidney disease
CC    associated with glomerulonephritis and enteric nervous system
CC    related disorders.  Transgenic and knockout animals are also
CC    claimed.
SQ    Sequence   840 BP;    205 A;    245 C;    200 G;    190 T;

  Query Match           35.0%;  Score 748;  DB 37;  Length 840;
  Best Local Similarity  94.5%;  Pred. No. 0.00e+00;
  Matches   794;  Conservative    0;  Mismatches  46;  Indels   0;  Gaps   0;

       1 cgccgcaagtgccacaaagccctcaggcagttcttcgacaaagttccagccaagcacagc 60
         ||||| |||||||||||| ||||||||||||||||||||| ||||| ||||||||||||
Qy    866 CGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGC 925

Db     61 tacgggatgctcttctgctcctgccgggacgtcgcctgcaccgagaggcggcgacagact 120
         |||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||
Qy    926 TACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCGGCGACAGACT 985

Db    121 atcgtccctgtgtgctcctatgaagaacgagagaggcccaactgcctgaatctgcaagac 180
         |||||||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||
Qy    986 ATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGAC 1045

Db    181 tcctgcaagacaaattacatctgcagatctcgccttgcagatttttttaccaactgccag 240
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1046 TCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATTTTTTTACCAACTGCCAG 1105

Db    241 ccagagtcaaggtctgtcagcaactgtcttaaggagaactacgcagactgcctcctggcc 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1106 CCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCC 1165

Db    301 tactcgggactgattggcacagtcatgactcctaactacatagactccagcagcctcagt 360
         |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||
Qy   1166 TACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAGACTCCAGCAGCCTCAGC 1225

Db    361 gtggcgccgtggtgcgattgcagcaacagtggcaatgacctggaagattgcctgaagttt 420
         ||||| || ||||| || |||||||||| |||||||||||||||||| ||| |||| |||
Qy   1226 GTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGGAAGACTGCTTGAAATTT 1285

Db    421 ctgaatttttttaaggacaatacgtgtctcaaaaatgcaattcaagcctttggcaatggc 480
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Qy   1286 CTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGC 1345

Db    481 tcggatgtgaccatgtggcagccagccccccccagtccagaccaccactgccacgactacc 540
         || ||||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Qy   1346 TCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCACCACTGCCACCACTACC 1405

Db    541 actgccttccggatcaagaacaagcctctagggccagcaggctctgagaatgagattccc 600
         |||||||||||| |||||||||||||||| |||||||||| |||||||||||||| |||
Qy   1406 ACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGTCTGAGAATGAGATCCCC 1465

Db    601 acacacgttttaccaccgtgtgctaatttgcaggcacagaagctgaaatccaatgtatcg 660
         |||||||||||||||| ||||| |||||||||| |||||||||||||||||||||| |||
Qy   1466 ACACACGTTTTACCACCCTGTGCGAATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCG 1525

Db    661 ggcagtacacatctctgtctttctgataatgattacggaaaggatggtctcgctggtgcc 720
         || || ||||| |||||||||||||||| ||||| ||||||||||||||||||||||||
Qy   1526 GGTAGCACACACCTCTGTCTTTCTGATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCC 1585

Db    721 tccagccacataaccacaaaatcaatggctgctcctcccagctgcggtctgagctcactg 780
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Qy   1586 TCCAGCCACATAACCACAAAATCAATGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTG 1645

Db    781 ccggtgatggtgttcaccgctctggctgccctgttgtctgtatcattggcagaaacatcg 840
         |||||| || || ||||||| || |||||||||| |||||||| |||||||||| |||
Qy   1646 CCGGTGCTGATGCTCACCGCCCTTGCTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCG 1705


RESULT    3
ID    T84976 standard; cDNA; 418 BP.
AC    T84976;
DT    27-APR-1998  (first entry)
DE    Mouse GDNFR alpha clone 26 5' end.
KW    Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW    GDNF; mouse; kidney disease; glomerulonephritis; therapy; ss.
OS    Mus musculus.
FH    Key             Location/Qualifiers
FT    CDS             287..418
FT                    /*tag= a
FT                    /note= "5' end of GDNFR alpha coding region"
PN    WO9733912-A2.
PD    18-SEP-1997.
PF    13-MAR-1997; U04363.
PR    14-MAR-1996; US-618236.
PR    14-MAR-1996; US-615902.
PA    (GETH ) GENENTECH INC.
PI    Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR    WPI; 97-470819/43.
PT    Isolated glial cell derived neurotrophic factor receptor alpha -
PT    useful to develop products to diagnose and treat associated
PT    disorders, particularly enteric nervous system or kidney disorders
PS    Example 1; Page 60; 100pp; English.
CC    This cDNA sequence comprises the 5' end of mouse full-length
CC    glial cell derived neurotrophic factor receptor alpha (GDNFR alpha)
CC    (see W27327) clone 26; the 3' end of the clone is given in
CC    T84977.  Clone 26 was isolated from a mouse cDNA library using
CC    rat GDNFR alpha cDNA (see T84975) as probe.  The invention relates
CC    to novel uses of GDNF and its receptor.  In particular, it relates
CC    to native rat GDNFR alpha (see W27327), its variants and soluble
CC    derivatives (extracellular domain), chimeric GDNFR alpha and
CC    antibodies which bind to the GDNFR alpha, including agonist and
CC    neutralising antibodies, as well as various uses for these
CC    molecules.  It also relates to assay systems for detecting ligands
CC    to GDNFR alpha, systems for studying the physiological role of
CC    GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC    methods for identifying molecules homologous to GDNFR alpha, and
CC    therapeutic techniques (claimed) for the treatment of GDNF-related
CC    and GDNFR alpha-related conditions, particularly kidney disease
```

```
CC   associated with glomerulonephritis and enteric nervous system
CC   related disorders.  Transgenic and knockout animals are also
CC   claimed.
SQ   Sequence   418 BP;    70 A;    119 C;    150 G;    79 T;

  Query Match           16.2%;  Score 347;  DB 37;  Length 418;
  Best Local Similarity  95.3%;  Pred. No. 2.00e-230;
  Matches   365;  Conservative    0;  Mismatches  18;  Indels   0;  Gaps   0;

Db     36 ggcaggttgggtcggaactgaacccctaaaagcgggtccgcctcccgccctcgcgcccgc 95
           |||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Qy     51 GGCAAGTTGGGTCGGAACTGAACCCCTGAAAGCGGGTCCGCCTCCCGCCCTCGCGCCCGC 110

Db     96 tcggagctgagtccctggcggcggtgggcggcagagcaacggggagtctgctctcaccct 155
           |||| ||||||| ||||||||||||||||||||||| ||||||||||||||||||||||
Qy    111 CCGGATCTGAGTCGCTGGCGGCGGTGGGCGGCAGAGCGACGGGGAGTCTGCTCTCACCCT 170

Db    156 ggatggagcttaactttgagtggccagaggagcgcagtcgcccggggatctctgcacgct 215
           |||||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||
Qy    171 GGATGGAGCTGAACTTTGAGTGGCCAGAGGAGCGCAGTCGCCCGGGGATCGCTGCACGCT 230

Db    216 gagctctctccgcgagatccggtggcggctttggatttgggggggcggggaccagctgc 275
           |||||||||| ||||| | || ||||||||||||||||||||||||||||||||||||
Qy    231 GAGCTCTCTCCCCGAGACCGGGCGGCGGCTTTGGATTTTGGGGGGGCGGGGACCAGCTGC 290

Db    276 gcggtggcaccatgttcctagccactctgtacttcgtgctgccactcctggatttgctga 335
           |||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Qy    291 GCGGCGGCACCATGTTCCTAGCCACTCTGTACTTCGCGCTGCCACTCCTGGATTTGCTGA 350

Db    336 tgtcggccgaggtgagtggtggggaccgcctggactgtgtgaaagccagtgatcagtgcc 395
           |||| |||||||||||||||| ||||| |||||||||||||||||| ||||||||||
Qy    351 TGTCCGCCGAGGTGAGTGGTGGAGACCGTCTGGACTGTGTGAAAGCCAGCGATCAGTGCC 410

Db    396 tgaaggaacagagctgcagcacc 418
           |||||||||||||||||||||||
Qy    411 TGAAGGAACAGAGCTGCAGCACC 433


RESULT    4
ID   T84979 standard; DNA; 453 BP.
AC   T84979;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence yI70a10.r1.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy;
KW   expressed sequence tag; EST; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a human EST-derived sequence designated
CC   yI70a10.r1.  It can be used to identify glial cell derived
CC   neurotrophic factor receptor (GDNFR) sequences, including human
CC   variants.  Also suitable as probes for GDNFR sequences are another
CC   human EST-derived sequence designated ye83h05.r1 (see T84978), and
CC   fragments of these EST-derived sequences (see T84980-81) or
CC   proteins encoded by them.  The invention relates to novel uses of
CC   GDNF and its receptor.  In particular, it relates to native rat
CC   GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
CC   Transgenic and knockout animals are also claimed.
SQ   Sequence   453 BP;    86 A;    152 C;    121 G;    91 T;

  Query Match            6.2%;  Score 133;  DB 37;  Length 453;
  Best Local Similarity  71.7%;  Pred. No. 1.41e-72;
  Matches   243;  Conservative    0;  Mismatches  93;  Indels   3;  Gaps   3;

Db      3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
          ||| | |||||||| || ||||||||||||||||| | |||| ||| ||||||    |
Qy    755 AACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGG 814

Db     63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
          ||  |||||||| ||  ||||| || | ||   | | ||| ||||  ||||||||||||
Qy    815 TCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCAACCGCCGT 871

Db    123 aagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
          |||||||||||||||||  | ||||||||||||  ||| ||   | || ||| |||| |
Qy    872 AAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGG 931

Db    183 atgctcttctgctcctgccaagaccaggcgtgcgctgagcgcgcgggcaaaaccatcctg 242
          |||||||||||||||||  |||   || ||| | ||||   | |  | || ||| |
Qy    932 ATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTC 991

Db    243 cccagctgctcctatgaggacaaggagaagcccaactgcctggacctgcgtggcgtgtgc 302
          |||   |||||||||| ||    |||| ||||||||||||   ||||  | |   |||
Qy    992 CCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGC 1051

Db    303 cggactgaccacctgtgtcggtcccggctngccgacttt 341
            |||  |  || | ||  | || || || || || |||
Qy   1052 AAGACCAATTACATCTGCAGATCTCGCCTTGCAGATTTT 1090


RESULT    5
ID   T84978 standard; DNA; 351 BP.
AC   T84978;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence ye83h05.r1.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW   expressed sequence tag; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a human EST-derived sequence designated
CC   ye83h05.r1.  It can be used to identify glial cell derived
CC   neurotrophic factor receptor (GDNFR) sequences, including human
CC   variants.  Also suitable as probes for GDNFR sequences are another
CC   human EST-derived sequence designated yI70a10.r1 (see T84979), and
CC   fragments of these EST-derived sequences (see T84980-81) or
CC   proteins encoded by them.  The invention relates to novel uses of
CC   GDNF and its receptor.  In particular, it relates to native rat
CC   GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
```

```
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
CC   Transgenic and knockout animals are also claimed.
SQ   Sequence   351 BP;    77 A;    117 C;    96 G;    57 T;

  Query Match            5.7%;  Score 122;  DB 37;  Length 351;
  Best Local Similarity  73.1%;  Pred. No. 1.01e-64;
  Matches   212;  Conservative    0;  Mismatches  75;  Indels   3;  Gaps   3;

Db     40 ccaagagcaaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaaga 99
           ||||  | ||| | |||||||| || ||||||||||||||||||| | |||| ||| ||||
Qy    747 CCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGA 806

Db    100 agctgcgctcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgca 159
           ||    | ||  |||||||| ||  ||||| || | ||   | | ||| |||| |||||
Qy    807 AGTACAGGTCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCA 863

      160 accgccgcaagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtaca 219
           ||||||| |||||||||||||||||  | ||||||||||||  ||| ||   | || |||
      864 ACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACA 923

Db    220 cctaccgcatgctcttctgctcctgccaagatcaggcgtgcgctgagcncgcggncaaaa 279
           |||| | |||||||||||||||||||  ||    || ||| | ||||    |    | |
Qy    924 GCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCGGCGACAGA 983

Db    280 ccatcctgcccagctgctcctatgaggacaaggagaagcccaactgcntg 329
           | ||| | |||   |||||||||| ||    |||| |||||||||| ||
Qy    984 CTATCGTCCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTG 1033


RESULT    6
ID   Q85526 standard; DNA; 385 BP.
AC   Q85526;
DT   10-JAN-1996  (first entry)
DE   DNA probe 32 detects DNA-protein complex in immortal cells.
KW   DNA-protein complex; detection; proliferation; tumour formation;
KW   diagnose; malignancy; biopsy; probe; ss.
OS   Synthetic.
PN   WO9502701-A1.
PD   26-JAN-1995.
PF   13-JUL-1994; E02307.
PR   15-JUL-1993; DE-323727.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Abken H,  Albert W,  Jungfer H,  Abken HJ;
DR   WPI; 95-067344/09.
PT   New DNA-protein complex characteristic of cells with unlimited
PT   proliferation capacity  - and its components and derived antibodies,
PT   useful in diagnosing malignant tumours.
PS   Claim 1; Page 73; 106pp; German.
CC   The DNA shown is found in human or animal cells that have an unlimited
CC   capacity for unlimited cell proliferation or tumour formation. They have
CC   no ability to promote immortalisation of the cells, and are usually
CC   found in a DNA-protein complex in the cell cytoplasm. The DNA is useful
CC   for detection of these complexes, and diagnosis of malignant tumours.
CC   Differentiation between cells with unlimited and only transient
CC   proliferation is possible. (See also Q85493-54)
SQ   Sequence   385 BP;    121 A;    83 C;    90 G;    91 T;

  Query Match            5.2%;  Score 112;  DB 15;  Length 385;
  Best Local Similarity 100.0%;  Pred. No. 1.25e-57;
  Matches   112;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 gatctcgccttgcagatttttttaccaactgccagccagagtcaaggtctgtcagcaact 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1071 GATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACT 1130

Db     61 gtcttaaggagaactacgcagactgcctcctggcctactcgggactgattgg 112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1131 GTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGG 1182


RESULT    7
ID   T84981 standard; DNA; 201 BP.
AC   T84981;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence yI70a10.r1 fragment.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW   expressed sequence tag; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a fragment of a human EST-derived
CC   sequence designated yI70a10.r1 (see T84979).  It can be used as
CC   a probe to identify glial cell derived neurotrophic factor receptor
CC   (GDNFR) sequences, including human variants.  The invention relates
CC   to novel uses of GDNF and its receptor.  In particular, it relates
CC   to rat GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
CC   Transgenic and knockout animals are also claimed.
SQ   Sequence   201 BP;    41 A;    78 C;    46 G;    36 T;

  Query Match            4.5%;  Score 96;  DB 37;  Length 201;
  Best Local Similarity  77.9%;  Pred. No. 2.10e-46;
  Matches   155;  Conservative    0;  Mismatches  41;  Indels   3;  Gaps   3;

Db      3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
          ||| | |||||||| || |||||||||||||||||| | |||| ||| ||||||    |
Qy    755 AACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGG 814

Db     63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
          ||  |||||||| ||  ||||| || | ||   | | ||| |||| ||||||||||||
Qy    815 TCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCAACCGCCGT 871

Db    123 aagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
          ||||||||||||||||  | ||||||||||||  ||| ||   | || ||| |||| |
Qy    872 AAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGG 931

Db    183 atgctcttctgctcctgcc 201
          ||||||||||||||||||||
Qy    932 ATGCTCTTCTGCTCCTGCC 950


RESULT    8
ID   T84980 standard; DNA; 201 BP.
AC   T84980;
DT   27-APR-1998  (first entry)
DE   Human EST-derived sequence ye83h05.r1 fragment.
KW   Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW   GDNF; human; kidney disease; glomerulonephritis; therapy; EST;
KW   expressed sequence tag; ss.
OS   Homo sapiens.
PN   WO9733912-A2.
PD   18-SEP-1997.
PF   13-MAR-1997; U04363.
```

```
PR   14-MAR-1996; US-618236.
PR   14-MAR-1996; US-615902.
PA   (GETH ) GENENTECH INC.
PI   Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR   WPI; 97-470819/43.
PT   Isolated glial cell derived neurotrophic factor receptor alpha -
PT   useful to develop products to diagnose and treat associated
PT   disorders, particularly enteric nervous system or kidney disorders
PS   Example 1; Page 61; 100pp; English.
CC   This DNA sequence comprises a fragment of a human EST-derived
CC   sequence designated ye83h05.r1 (see T84978).  It can be used as
CC   a probe to identify glial cell derived neurotrophic factor receptor
CC   (GDNFR) sequences, including human variants.  The invention relates
CC   to novel uses of GDNF and its receptor.  In particular, it relates
CC   to rat GDNFR alpha (see W27327), its variants and soluble derivatives
CC   (extracellular domain), chimeric GDNFR alpha and antibodies which
CC   bind to the GDNFR alpha, including agonist and neutralising
CC   antiodies, as well as various uses for these molecules.  It also
CC   relates to assay systems for detecting ligands to GDNFR alpha,
CC   systems for studying the physiological role of GDNF, diagnostic
CC   techniques for identifying GDNF-related conditions, methods for
CC   identifying molecules homologous to GDNFR alpha, and claimed
CC   methods for the treatment of GDNF-related and GDNFR alpha-related
CC   conditions, particularly kidney disease associated with
CC   glomerulonephritis and enteric nervous system related disorders.
CC   Transgenic and knockout animals are also claimed.
SQ   Sequence   201 BP;    41 A;    78 C;    46 G;    36 T;

  Query Match             4.5%;  Score 96;  DB 37;  Length 201;
  Best Local Similarity  77.9%;  Pred. No. 2.10e-46;
  Matches   155;  Conservative    0;  Mismatches  41;  Indels   3;  Gaps   3;

Db      3 aaccattgcctggatgctgccaaggcctgcaacctgaatgacaactgcaagaagctgcgc 62
          ||| | |||||||| || ||||||||||||||||||| | |||| ||| ||||||    |
Qy    755 AACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAGG 814

Db     63 tcctcctacatctccatctgcaaccgcgagatctcgcccaccgagcgctgcaaccgccgc 122
          ||  |||||||| ||  ||||| || | ||   | | ||| ||||  ||||||||||||
Qy    815 TCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCAACCGCCGT 871

Db    123 aagtgccacaaggccctgcgccagttcttcgaccgggtgcccagcgagtacacctaccgc 182
          ||||||||||||||||  | ||||||||||||  ||| ||   | || ||| |||| |
Qy    872 AAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGG 931

Db    183 atgctcttctgctcctgcc 201
          ||||||||||||||||||
Qy    932 ATGCTCTTCTGCTCCTGCC 950


RESULT    9
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             2.1%;  Score 44;  DB 9;  Length 91;
  Best Local Similarity   9.3%;  Pred. No. 7.63e-12;
  Matches     5;  Conservative   44;  Mismatches   5;  Indels   0;  Gaps   0;

Db     13 vhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctcaag 66
          ::::: :::: :::::::::::::::::::: :::::::::: :::: |||||
Qy   1370 GCCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCTTCCGGGTCAAG 1423


RESULT   10
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match             2.1%;  Score 44;  DB 9;  Length 91;
  Best Local Similarity   8.9%;  Pred. No. 7.63e-12;
  Matches     5;  Conservative   45;  Mismatches   6;  Indels   0;  Gaps   0;

Db      5 ccggcgssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
          ||| ||:::::::::::::: :::::::::: ::::: :::::::::  ::::::
Cp    297 CCGCCGCGCAGCTGGTCCCCGCCCCCCCAAAATCCAAAGCCGCCGCCCGGTCTCGG 242


RESULT   11
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
```

```
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            2.0%;  Score 43;  DB 1;  Length 204;
  Best Local Similarity  22.1%;  Pred. No. 3.11e-11;
  Matches    36;  Conservative   62;  Mismatches  64;  Indels   1;  Gaps   1;

Db     23 gcyyrcayrccbgcaggycgacbcyrraggnyccccgggywccgagcycgaayycdchvg 82
          ||::  | :||: || | ||   | : |     ||  ||:  | || :|| |:  :|::|
Cp    542 GCTTCAAGGCCTCCATGGCGCTACGGCACTCATCCTTGGCCTCAAGGCCGGATG-TCAGG 484

       83 ccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbv 142
          | | ::||  ::  : : :: :: ::: | : ||: :: :::|   ||::::: :::: :
      483 CTGAAGTTGGTTTCCTTGCCCGCCACGCACTGCCTTAGTGTCGCGTACTTGGTGCTGCAG 424

Db    143 ynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvh 185
          : :: :  |  | :  ::    ::   : : ::::: ::: ::
Cp    423 CTCTGTTCCTTCAGGCACTGATCGCTGGCTTTCACACAGTCCA 381


RESULT   12
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match            1.9%;  Score 40;  DB 1;  Length 204;
  Best Local Similarity  11.7%;  Pred. No. 1.99e-09;
  Matches    12;  Conservative   55;  Mismatches  35;  Indels   1;  Gaps   1;

Db     85 gymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyv-ybbbvy 143
          |::: ||:::::::  :::: : :| |::|| :::::: ||   |:::: : :::: :
Qy    667 GCCGGTTAACAGCAGGTTGTCAGATATATTCCGGGCAGTCCCGTTCATATCAGATGTTTT 726

Db    144 nvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          :: :          :: : ::    : :|:::::   ::::|
Qy    727 CCAGCAAGTGGAACACATTTCCAAAGGGAACAACTGCCTGGAC 769


RESULT   13
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match            1.8%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 3.05e-08;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnn 62
          :  :  :  :  :  :  :  :  :  :|||  :  :  :  :  :  :  :
Qy    844 CATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGA 903

Db     63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :  :  :  :  :  :  :||  :  :  :  :  :  :  :  :  :
Qy    904 CAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGAC 955


RESULT   14
```

```
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
     30-DEC-1993; US-176500.
     31-JAN-1994; US-189331.
     (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match             1.8%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   4.5%;  Pred. No. 3.05e-08;
  Matches     5;  Conservative   34;  Mismatches  73;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  :|||  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy    844 CATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGA 903

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|| |  :  :  :
Qy    904 CAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCGGGAC 955


RESULT   15
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match             1.8%;  Score 38;  DB 12;  Length 114;
  Best Local Similarity   7.1%;  Pred. No. 3.05e-08;
  Matches     8;  Conservative   32;  Mismatches  72;  Indels   0;  Gaps   0;

Db      1 tgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnn 60
            |||  :  :  :  :  :  :  :  :  :  :  :||  :  :  :  :  :  :
Qy    833 TGCACCACCAGCATGTCCAACGAGGTCTGCAACCGCCGTAAGTGCCACAAGGCCCTCAGG 892

Db     61 nnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbt 112
             :  :| |  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy    893 CAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCT 944


Search completed: Wed Jun 24 23:20:17 1998
Job time : 600 secs.
```

****

****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Wed Jun 24 22:01:59 1998; MasPar time 2326.45 Seconds
1227.537 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-3
Description: (1-2138) from US08866354.seq
Perfect Score: 2138
N.A. Sequence: 1 AGCTCGCTCTCCCGGGGCAG.....AGACAGGCCCGCAGCCGTCG 2138
Comp: TCGAGCGAGAGGGCCCCGTC.....TCTGTCCGGGCGTCGGCAGC

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1759237 seqs, 667866413 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est54
1:em_est1 2:em_est3
Database: genbank-est106
3:gb_est1 4:gb_est10 5:gb_est11 6:gb_est12 7:gb_est13
8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17
12:gb_est18 13:gb_est19 14:gb_est2 15:gb_est20
16:gb_est21 17:gb_est22 18:gb_est3 19:gb_est4 20:gb_est5
21:gb_est6 22:gb_est7 23:gb_est8 24:gb_est9 25:gb_gss
26:gb_sts

Statistics: Mean 11.800; Variance 1.984; scale 5.947

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	496	23.2	631	7	AA518362	vi10d02.r1 Barstead mo	0.00e+00
	2	408	19.1	446	5	AA387098	vc18b07.r1 Ko mouse em	0.00e+00
	3	401	18.8	465	21	W53498	md45e01.r1 Soares mous	0.00e+00
	4	378	17.7	430	21	W76716	me81h06.r1 Soares mous	0.00e+00
	5	313	14.6	478	24	AA250120	mx21g08.r1 Soares mous	0.00e+00
c	6	303	14.2	515	7	AA512935	nh91b04.s1 NCI_CGAP_Br	0.00e+00
	7	220	10.3	439	12	AA727460	vu96d10.r1 Stratagene	0.00e+00
	8	160	7.5	459	24	AA245689	mx03e07.r1 Soares mous	6.37e-300
c	9	149	7.0	330	9	AA573164	nm51f06.s1 NCI_CGAP_Br	1.20e-273
	10	141	6.6	266	7	AA471183	PMY2139 KG1a Lambda Za	1.21e-254
	11	119	5.6	521	18	H12981	yl70a10.r1 Homo sapien	6.57e-203
	12	100	4.7	229	3	R02249	ye83h05.r1 Homo sapien	4.72e-159
c	13	62	2.9	252	13	AA754459	97SN1787 Rice Immature	3.14e-75
	14	60	2.8	252	13	AA754459	97SN1787 Rice Immature	4.79e-71
c	15	54	2.5	543	24	AA270188	vb11e06.r1 Soares mous	1.06e-58
c	16	52	2.4	247	13	AA754458	97SN1784 Rice Immature	1.15e-54
	17	49	2.3	247	13	AA754458	97SN1784 Rice Immature	1.08e-48
c	18	45	2.1	180	23	AA064450	ml47h11.r1 Stratagene	6.55e-41
	19	36	1.7	284	25	FR0012397	F.rubripes GSS sequenc	2.50e-24
	20	36	1.7	2275	12	AF034173	Homo sapiens ntcon2 co	2.50e-24
	21	35	1.6	350	4	AA349976	EST56965 Infant brain	1.38e-22
	22	29	1.4	478	22	W73681	zd55h01.r1 Soares feta	1.08e-12
c	23	29	1.4	569	25	FR0020899	F.rubripes GSS sequenc	1.08e-12
c	24	30	1.4	596	25	FR0022657	F.rubripes GSS sequenc	2.88e-14
c	25	31	1.4	2275	12	AF034173	Homo sapiens ntcon2 co	7.13e-16
	26	27	1.3	369	19	N23527	yv91f01.r1 Homo sapien	1.18e-09
c	27	27	1.3	370	25	AF046247	Mus musculus clone OST	1.18e-09
c	28	27	1.3	416	18	M78296	EST00444 Homo sapiens	1.18e-09
	29	28	1.3	466	13	AA778124	zf46g09.s1 Soares feta	3.73e-11
	30	27	1.3	553	13	AA568016	HL02152.5prime HL Dros	1.18e-09
c	31	26	1.2	193	25	FR0012700	F.rubripes GSS sequenc	3.42e-08
c	32	26	1.2	202	25	FR0014400	F.rubripes GSS sequenc	3.42e-08
c	33	26	1.2	231	5	W78691	EST00010 Tf cDNA libra	3.42e-08
c	34	26	1.2	301	23	AA218487	MAAD0155.M3R Schistoso	3.42e-08
c	35	26	1.2	304	23	AA218491	MAAD0158.M3R Schistoso	3.42e-08
c	36	26	1.2	304	25	FR0015974	F.rubripes GSS sequenc	3.42e-08
c	37	25	1.2	326	26	HUMUT8005B	Human STS UT8005, 3' p	8.97e-07
	38	26	1.2	344	26	HS299YF1	H. sapiens (D18S476) D	3.42e-08
c	39	26	1.2	373	7	AA525583	MAAD0146.RAR Schistoso	3.42e-08
	40	26	1.2	382	6	HUM356C04B	Human aorta cDNA 5'-en	3.42e-08
c	41	26	1.2	420	13	AA754253	97MJ0074 Rice Immature	3.42e-08
c	42	26	1.2	544	24	C22979	Dictyostelium discoide	3.42e-08
c	43	26	1.2	547	12	W43683	23300 CD4-16 Arabidops	3.42e-08
	44	26	1.2	668	5	W78683	EST00013 Tf cDNA libra	3.42e-08
c	45	26	1.2	1287	13	AF038250	Homo sapiens clone ntc	3.42e-08

ALIGNMENTS

RESULT 1
LOCUS AA518362 631 bp mRNA EST 16-JUL-1997
DEFINITION vi10d02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone 903363 5', mRNA sequence.
ACCESSION AA518362
NID g2259047
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 631)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:524027
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 259.
FEATURES Location/Qualifiers
source 1. .631
/organism="Mus musculus"
/strain="C3H"

```
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                3']; double-stranded cDNA was ligated to Eco RI adaptors
                [AATTCGGATCCTTG], digested with Not I and cloned into the
                Not I and Eco RI sites of the modified pT7T3 vector.
                Library constructed by Bob Barstead. The C2C12 cell line
                (available from ATCC, catalog # CRL-1772) differentiates
                rapidly, forming contractile myotubes and producing
                characteristic muscle proteins."
                /db_xref="taxon:10090"
                /clone="903363"
                /clone_lib="Barstead mouse myotubes MPLRB5"
                /cell_line="C2C12"
                /lab_host="DH10B"
     mRNA       <1. .>631
BASE COUNT      154 a    174 c    166 g    133 t      4 others
ORIGIN

Query Match           23.2%;  Score 496;  DB 7;  Length 631;
Best Local Similarity  93.4%;  Pred. No. 0.00e+00;
Matches   550;  Conservative    0;  Mismatches  36;  Indels   3;  Gaps   3;

Db      14 GTGCCACAAAGCCCTCAGGCAGTTCTTCGACAAAGTACCAGCCAAGCACAGCTACGGGAT 73
           ||||||||| |||||||||||||||||||||||| || || ||||||||||||||||||||
Qy     874 GTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGAT 933

Db      74 GCTCTTCTGCTCCTGCCGGGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCC 133
           ||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Qy     934 GCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCC 993

Db     134 TGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAA 193
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Qy     994 CGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAA 1053

Db     194 GACAAATTACATCTGCAGATCTCGCCTTGCAGAGTTTTTTTACCAACTGCCAGCCAGAGT 253
           ||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Qy    1054 GACCAATTACATCTGCAGATCTCGCCTTGCAGA-TTTTTTTACCAACTGCCAGCCAGAGT 1112

Db     254 CAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGG 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1113 CAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGG 1172

Db     314 GACTGATTGGCACAGTCATGACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGC 373
           |||||||||||||||||||||||| |||||| |||||||||||||||||| ||||| |
Qy    1173 GACTGATTGGCACAGTCATGACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCAC 1232

      374 CGTGGTGCGATTGCAGCAGCAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATT 433
           | ||||| || ||||||| ||| |||||||||||||||| ||| |||| ||||||||||
Qy    1233 CATGGTGTGACTGCAGCAACAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATT 1292

Db     434 TTTTTAAGGGCAATACGTGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCGGATG 493
           ||||||||| |||||| ||||||||||||||||||||||||||||||||||||| ||||
Qy    1293 TTTTTAAGGACAATACTTGTCTCAAAAATGCAATTCAAGCCTTTGGCAATGGCTCAGATG 1352

Db     494 TGACCATGTGGCAGCCAGCACCAC-AGTCCAGACCACCAATGCCACGACTACCAATGCCT 552
           |||||||||||||||||| || | |||||||||||||| |||||| ||||||| |||||
Qy    1353 TGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCACCACTGCCACCACTACCACTGCCT 1412

Db     553 TCGGGATCAAGAACGAGCTACTGGG-CCGGCAGGTTCTGAGAATGAGAT 600
           || || |||||||| ||| |||| || ||||| ||||||||||||||
Qy    1413 TCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGTCTGAGAATGAGAT 1461


RESULT    2
LOCUS       AA387098      446 bp    mRNA            EST       23-APR-1997
DEFINITION  vc18b07.r1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone 774901
            5', mRNA sequence.
ACCESSION   AA387098
NID         g2040052
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:467757
            High quality sequence stop: 413.
FEATURES              Location/Qualifiers
     source           1. .446
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total
                      RNAs were extracted from 11.5 dpc embryos (excluding
                      placenta and yolk sac). The double-stranded cDNA was
                      synthesized with an oligo (dT)-1 primer
                      GAGAGAGACTAGTTCTAGATCGCGAGCGGCCGCTTTTTTTTTTTTTTTT 3'.
                      The cDNAs were ligated to LL-Sal3A: 5'
                      GCTATTGACGTCGACTATCC 3' and LL-Sal3B: 5'
                      GGATAGTCGACGTCAAT 3'. The cDNAs were size-selected and
                      amplified by long-range PCR using Ex Taq polymerase for 18
                      cycles. The PCR-amplifiable cDNA mixture went through
                      one round of equalization and was digested with SalI/NotI
                      and cloned into the SalI/NotI sites of the pSPORT1
                      plasmid vector (Life Technologies). The library was
                      constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong
                      Wang."
                      /db_xref="taxon:10090"
                      /clone="774901"
                      /clone_lib="Ko mouse embryo 11 5dpc"
                      /sex="pooled"
                      /dev_stage="11.5dpc"
                      /lab_host="DH10B"
     mRNA             <1. .>446
BASE COUNT      112 a    122 c    107 g    105 t
ORIGIN

  Query Match           19.1%;  Score 408;  DB 5;  Length 446;
  Best Local Similarity  95.7%;  Pred. No. 0.00e+00;
  Matches   427;  Conservative    0;  Mismatches  19;  Indels   0;  Gaps   0;

Db       1 GCAGTTCTTCGACAAAGTTCCAGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 60
           ||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Qy     892 GCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGGGATGCTCTTCTGCTCCTGCCG 951

Db      61 GGACGTCGCCTGCACCGAGAGGCGGCGACAGACTATCGTCCCTGTGTGCTCCTATGAAGA 120
           |||| ||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Qy     952 GGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGTCCCCGTGTGCTCCTATGAAGA 1011

Db     121 ACGAGAGAGGCCCAACTGCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAG 180
           |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Qy    1012 ACGAGAGAGGCCCAACTGCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAG 1071

Db     181 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1072 ATCTCGCCTTGCAGATTTTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTG 1131
```

```
Db     241 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1132 TCTTAAGGAGAACTACGCAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCAT 1191

Db     301 GACTCCTAACTACATAGACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAA 360
           |||||| |||||| |||||||||||||||||| ||||| || ||||| || ||||||||
Qy    1192 GACTCCCAACTACGTAGACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAA 1251

Db     361 CAGTGGCAATGACCTGGAAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTG 420
           ||| |||||||||||||||| ||| |||| ||||||||||||||||||||||||||| ||
Qy    1252 CAGCGGCAATGACCTGGAAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTG 1311

Db     421 TCTCAAAAATGCAATTCAAGCCTTTG 446
           |||||||||||||||||||||||||
Qy    1312 TCTCAAAAATGCAATTCAAGCCTTTG 1337
```

```
RESULT    3
LOCUS       W53498         465 bp    mRNA            EST       03-JUN-1996
DEFINITION  md45e01.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 371352 5', mRNA sequence.
ACCESSION   W53498
NID         g1357323
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:232784
            Seq primer: mob.REGA+ET
            High quality sequence stop: 347.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="371352"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
```

```
     mRNA            <1. .>465
BASE COUNT      120 a    133 c    103 g    109 t
ORIGIN

  Query Match           18.8%;  Score 401;  DB 21;  Length 465;
  Best Local Similarity  94.0%;  Pred. No. 0.00e+00;
  Matches   439;  Conservative    0;  Mismatches  26;  Indels   2;  Gaps   2;

Db       1 GCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATT 60
           |||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Qy    1029 GCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATT 1088

Db      61 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1089 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 1148

Db     121 CAGACTGCCTCCTG-CCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAG 179
           |||||||||||||| |||||||||||||||||||||||||||||||| |||||| |||
Qy    1149 CAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAG 1208

Db     180 ACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGG 239
           |||||||||||||||| ||||| || ||||| || |||||||||| ||||||||||||
Qy    1209 ACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGG 1268

Db     240 AAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGACAATACGTGTCTCAAAAATGCAATTC 299
           |||| ||| |||| |||||||||||||||||||||||||||| ||||||||||||||||||
Qy    1269 AAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTC 1328

Db     300 AAGCCTTTGCACATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCAGTCCAGACCA 359
           |||||||||   ||||||| |||||||||||||||||||||||| ||||||||||||||
Qy    1329 AAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCA 1388

Db     360 CCACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCT 419
           |||||||||| |||||||||||||||||| |||||||||||||||| |||||||||||| |
Qy    1389 CCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGT 1448

Db     420 CTGAGAATGAGATTCC-ACACACGTTTTACCACCGTGTGCTAATTTG 465
           |||||||||||||| || |||||||||||||||||| ||||| ||||||
Qy    1449 CTGAGAATGAGATCCCCACACACGTTTTACCACCCTGTGCGAATTTG 1495
```

```
RESULT    4
LOCUS       W76716         430 bp    mRNA            EST       20-JUN-1996
DEFINITION  me81h06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone 402011 5', mRNA sequence.
ACCESSION   W76716
NID         g1387762
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:245779
            Seq primer: mob.REGA+ET
```

```
          High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="402011"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
     mRNA            <1. .>430
BASE COUNT      112 a    121 c    100 g     97 t
ORIGIN

  Query Match           17.7%;  Score 378;  DB 21;  Length 430;
  Best Local Similarity  94.4%;  Pred. No. 0.00e+00;
  Matches   407;  Conservative    0;  Mismatches  23;  Indels   1;  Gaps   1;

Db      1 GCCTGAATCTGCAAGACTCCTGCAAGACAAATTACATCTGCAGATCTCGCCTTGCAGATT 60
          |||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Qy   1029 GCCTGAGTCTGCAAGACTCCTGCAAGACCAATTACATCTGCAGATCTCGCCTTGCAGATT 1088

Db     61 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1089 TTTTTACCAACTGCCAGCCAGAGTCAAGGTCTGTCAGCAACTGTCTTAAGGAGAACTACG 1148

Db    121 CAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCTAACTACATAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||
Qy   1149 CAGACTGCCTCCTGGCCTACTCGGGACTGATTGGCACAGTCATGACTCCCAACTACGTAG 1208

Db    181 ACTCCAGCAGCCTCAGTGTGGCGCCGTGGTGCGATTGCAGCAACAGTGGCAATGACCTGG 240
          |||||||||||||||| ||||| || ||||| || |||||||||| ||||||||||||
Qy   1209 ACTCCAGCAGCCTCAGCGTGGCACCATGGTGTGACTGCAGCAACAGCGGCAATGACCTGG 1268

      241 AAGATTGCCTGAAGTTTCTGAATTTTTTTAAGGA-AATACGTGTCTCAAAAATGCAATTC 299
          |||| ||| |||| |||||||||||||||||||| ||||| ||||||||||||||||||
Qy   1269 AAGACTGCTTGAAATTTCTGAATTTTTTTAAGGACAATACTTGTCTCAAAAATGCAATTC 1328

Db    300 AAGCCTTTGCAGATGGCTCGGATGTGACCATGTGGCAGCCAGCCCCCCCCAGTCCAGACCA 359
          |||||||||   ||||||| |||||||||||||||||||||||||| ||||||||||||
Qy   1329 AAGCCTTTGGCAATGGCTCAGATGTGACCATGTGGCAGCCAGCCCCTCCAGTCCAGACCA 1388

Db    360 CCACTGCCACGACTACCACTGCCTTCCGGATCAAGAACAAGCCTCTAGGGCCAGCAGGCT 419
          |||||||||| |||||||||||||||||| |||||||||||||||| |||||||||| |
Qy   1389 CCACTGCCACCACTACCACTGCCTTCCGGGTCAAGAACAAGCCTCTGGGGCCAGCAGGGT 1448

Db    420 CTGAGAATGAG 430
          |||||||||||
Qy   1449 CTGAGAATGAG 1459


RESULT    5
LOCUS       AA250120      478 bp    mRNA            EST       12-MAR-1997
DEFINITION  mx21g08.r1 Soares mouse NML Mus musculus cDNA clone 680894 5', mRNA
            sequence.
ACCESSION   AA250120
NID         g1882418
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:420598
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="680894"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
     mRNA            <1. .>478
BASE COUNT      121 a    116 c    111 g    130 t
ORIGIN

  Query Match           14.6%;  Score 313;  DB 24;  Length 478;
  Best Local Similarity  89.3%;  Pred. No. 0.00e+00;
  Matches   424;  Conservative    0;  Mismatches  39;  Indels  12;  Gaps   4;

Db      1 GCTAATTTGCAGGCACAGAAGCTGAAATCCAATGTATCGGGCAGTACACATCTCTGTCTT 60
          || |||||||||| |||||||||||||||||||| ||||| || ||||| |||||||||
Qy   1487 GCGAATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTCTT 1546

Db     61 TCTGATAATGATTACGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAA 120
          ||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Qy   1547 TCTGATAGTGATTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAA 1606

Db    121 TCAATGGCTGCTCCTCCCAGCTGCGGTCTGAGCTCACTGCCGGTGATGGTGTTCACCGCT 180
          |||||||||||||||||||||||| |||||||||||||||||| || || |||||||
Qy   1607 TCAATGGCTGCTCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCC 1666

Db    181 CTGGCTGCCCTGTTGTCTGTATCATTGGCAGAAACATCGTAGCTGCATCC--------AG 232
          || |||||||||| |||||||| |||||||||| |||||||||||||||        ||
Qy   1667 CTTGCTGCCCTGTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAG 1726

Db    233 TATGAAAAGACAAAAGAAAACCAAGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAAT 292
          ||||||||||||||||| |||||||||||||||||| ||||||||||||||||||||||
Qy   1727 TATGAAAAGACAAAAGAGAACCAAGTATTCTGTCCC-TGTCCTCTTGTATATCTGAAAAT 1785

Db    293 CCAGTTTTAAAAGCTCGGTTGAGAAGCAGTTTCACCCAGCGGGAACTCCTTTGTTGTTTT 352
          |||||||||||||| |||||||||||||||||||| | ||||||| ||  |||||||
Qy   1786 CCAGTTTTAAAAGCTCCGTTGAGAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTT 1845
```

```
Db    353 TAAGAAAGCTTTTTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTC 412
          |||||||||||  | |  |||||||  |||||||||||||||||||||||||||||||||
Qy   1846 TAAGAAAGCTTGT-G--GCCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTC 1902

Db    413 CAAACTCAGAAAACTCTGGGGCATGGTGTGGCTTCAGGGGACTGTTTGTAACCTG 467
          ||||| || ||  |||||||| ||||| ||||| |||||||  |||| | | ||
Qy   1903 CAAACCCATAAGGCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATG 1957


RESULT    6
LOCUS       AA512935      515 bp    mRNA            EST       05-AUG-1997
DEFINITION  nh91b04.s1 NCI_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:965839,
            mRNA sequence.
ACCESSION   AA512935
NID         g2251358
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1735   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 460.
FEATURES              Location/Qualifiers
     source           1. .515
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from pooled bulk
                      breast tumor tissue, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is not normalized. (The normalized
                      version of this library is NCI_CGAP_Br2.) Library was
                      constructed by Bento Soares and M. Fatima Bonaldo."
                      /db_xref="taxon:9606"
                      /clone="IMAGE:965839"
                      /clone_lib="NCI_CGAP_Br1.1"
                      /sex="female, pooled"
                      /tissue_type="breast"
                      /lab_host="DH10B"
     mRNA             <1. .>515
BASE COUNT      123 a    112 c    131 g    149 t
ORIGIN

  Query Match           14.2%;  Score 303;  DB 7;  Length 515;
  Best Local Similarity  85.0%;  Pred. No. 0.00e+00;
  Matches   398;  Conservative    0;  Mismatches  65;  Indels   5;  Gaps   3;

Db     53 TAAAGATAATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACA 112
          || |||||| ||||  |  || ||||| | || ||| ||| |||||| ||||||||  ||
Cp   1687 TACAGATAACAGGGCAGCAAGGGCGGTGAGCATCAGCACCGGCAGTGAGCTCAGACTGCA 1628

Db    113 GCTTGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAA-CACC---GAGACCTTC 168
          ||| ||||||||||||||||||||||||||||||||||||||||  ||||   |||||| ||
Cp   1627 GCTGGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAGGCACCAGCGAGACCATC 1568

Db    169 TTTTTCATAATTACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAG 228
          ||| |  ||| || ||  |||| ||||||||||||  | ||||||||||||||||||||
Cp   1567 CTTTCCGAAATCACTATCAGAAAGACAGAGGTGTGTGCTACCCGACACATTGGATTTCAG 1508

Db    229 CTTCTGTGCCTGTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGA 288
          |||||| ||||| ||||| ||||| ||||| ||||| || ||||| || ||||||||||||
Cp   1507 CTTCTGAGCCTGCAAATTCGCACAGGGTGGTAAAACGTGTGTGGGGATCTCATTCTCAGA 1448

Db    289 CCCTGCTGGCCCCAGGGCTCTGTTCTTAACCCGGA-GGCAGTGGTGGTAGCGGCAGTGGT 347
          ||||||||||||||  |   ||||||| ||||||| ||||||||| || | |||||||||
Cp   1447 CCCTGCTGGCCCCAGAGGCTTGTTCTTGACCCGGAAGGCAGTGGTAGTGGTGGCAGTGGT 1388

Db    348 GGTCTGTACTGGGAAGGCTGGCTGCCACACGGTCACATCGGAGCCATTGCCAAAGGCTTG 407
          |||||| |||||   |||||||||||||| |||||||| ||||||||||||||||||||
Cp   1387 GGTCTGGACTGGAGGGGCTGGCTGCCACATGGTCACATCTGAGCCATTGCCAAAGGCTTG 1328

Db    408 AATTGCATTTTTAAGACATGTATTGTCCTTGAAGAAATTCAAAAATTTCAAGCACTCTTC 467
          |||||||||||| ||||| |||||||||||  || ||||||| |||||||||||| |||||
Cp   1327 AATTGCATTTTTGAGACAAGTATTGTCCTTAAAAAAATTCAGAAATTTCAAGCAGTCTTC 1268

Db    468 TAGGTCGTTCCCACTGTTGCTGCAGTCACACCATGGGGCCACACTGAG 515
          ||||| || || ||||||||||||||||||||||| ||||| |||||
Cp   1267 CAGGTCATTGCCGCTGTTGCTGCAGTCACACCATGGTGCCACGCTGAG 1220


RESULT    7
LOCUS       AA727460      439 bp    mRNA            EST       02-JAN-1998
DEFINITION  vu96d10.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1210003 5', mRNA sequence.
ACCESSION   AA727460
NID         g2745167
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:646347
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 436.
FEATURES              Location/Qualifiers
     source           1. .439
                      /organism="Mus musculus"
                      /strain="C57BL/6"
                      /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="taxon:10090"
                      /clone="1210003"
```

```
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       105 a     79 c    117 g    138 t
ORIGIN

  Query Match           10.3%;  Score 220;  DB 12;  Length 439;
  Best Local Similarity  83.2%;  Pred. No. 0.00e+00;
  Matches   326;  Conservative    0;  Mismatches  58;  Indels   8;  Gaps   6;

Db       1 AGTATTCTGTCCCCTGTCCTCTTGTATATCTGAAAATCCAGGTTTTAAAAGCTCGGTTGA 60
          |||||||||||||  |||||||||||||||||||||||||||||| ||||||||||| |||||
Qy    1750 AGTATTCTGTCCC-TGTCCTCTTGTATATCTGAAAATCCAG-TTTTAAAAGCTCCGTTGA 1807

Db      61 GAAGCAGTTTCACCCAGC-GGAACTCCTTTGTTGTTTTTAAGAAAGCTTTTTGCCGCCCT 119
          |||||||||||||||| | ||||||| ||  |||||||||||||||||| | |  |||||
      1808 GAAGCAGTTTCACCCAACTGGAACTCTTTCCTTGTTTTTAAGAAAGCTTGT-G--GCCCT 1864

       120 CACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACTCAGAAAACTCTGGGGC 179
           ||  ||||||||||||||||||||||||||||||||||||||||| || ||  |||||||||
Qy    1865 CAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCATAAGGCTCTGGGGC 1924

Db     180 ATGGTGTGGCTTCAGGGGACTGTTTGTAACCTGGACTGTAAAGCAAGCTGAGCTTATGTT 239
           |||||  |||||  |||||||  |||| | | || | | || |||  | |  ||||
Qy    1925 GTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGGCTTATC--ATGTG 1982

Db     240 TTTGATGGTGATGGTGGTGGTGATGATGAAGATGATAGTTTTAACAGCTTGAACCCTGTT 299
           |||||||||||  | ||||  |||  ||||||  ||||  ||  ||||||||||||||||||||
Qy    1983 TTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCTTGAACCCTGTT 2042

Db     300 CTCTCTACTGGCTAGGAGGAGGAGATGCTATTGATAGAGATTCTTCCGGATCTTATTTAA 359
           ||||||||||| |||||  |||||||  |||||||||  ||||||||||   |||||  |  |
Qy    2043 CTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCATGTCTTACTCAG 2102

Db     360 TAGCGTTGGATTCTAAAGACATGACTGCAGCC 391
           |||  |||  ||||  ||||||  |  |  ||||||
Qy    2103 CAGCATTGCCTTCTGAAGACAGGCCCGCAGCC 2134


RESULT    8
LOCUS       AA245689      459 bp    mRNA            EST       10-MAR-1997
DEFINITION  mx03e07.r1 Soares mouse NML Mus musculus cDNA clone 679140 5', mRNA
            sequence.
ACCESSION   AA245689
NID         g1876478
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:418844
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="679140"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
     mRNA            <1. .>459
BASE COUNT       117 a     87 c    118 g    137 t
ORIGIN

  Query Match           7.5%;  Score 160;  DB 24;  Length 459;
  Best Local Similarity  80.1%;  Pred. No. 6.37e-300;
  Matches   226;  Conservative    0;  Mismatches  54;  Indels   2;  Gaps   1;

Db      13 TTGCCGCCCTCACAGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACTCAGAAA 72
           |||  |||||||  ||||||||||||||||||||||||||||||||||||||||| || ||
Qy    1855 TTGTGGCCCTCAGGGGCTTCTGTTGAAGAACTGCTACAGGGCTAATTCCAAACCCATAAG 1914

Db      73 ACTCTGGGGCATGGTGTGGCTTCAGGGGACTGTTTGTAACCTGGACTGTAAAGCAAGCTG 132
           |||||||||  |||||  |||||  |||||||  |||| | | || |  | ||     |||
Qy    1915 GCTCTGGGGCGTGGTGCGGCTTAAGGGGACCATTTGCACCATGTAAAGCAAGCTGGGCTT 1974

Db     133 AGCTTATGTTTTTGATGGTGATGGTGGTGGTGATGATGAAGATGATAGTTTTAACAGCTT 192
           | |  |||| ||||||||||| | |||| ||| |||||| |||| || ||||||||||||
Qy    1975 ATC--ATGTGTTTGATGGTGAGGATGGTAGTGGTGATGATGATGGTAATTTTAACAGCTT 2032

Db     193 GAACCCTGTTCTCTCTACTGGCTAGGAGGAGGAGATGCTATTGATAGAGATTCTTCCGGA 252
           |||||||||||||||||||| |||||  |||||||  |||||||| ||||||||||
Qy    2033 GAACCCTGTTCTCTCTACTGGTTAGGAACAGGAGATACTATTGATAAAGATTCTTCCATG 2092

Db     253 TCTTATTTAATAGCGTTGGATTCTAAAGACATGACTGCAGCC 294
           |||||  |  |  |||  |||  ||||  ||||||  |  |  ||||||
Qy    2093 TCTTACTCAGCAGCATTGCCTTCTGAAGACAGGCCCGCAGCC 2134


RESULT    9
LOCUS       AA573164      330 bp    mRNA            EST       28-AUG-1997
DEFINITION  nm51f06.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1071779, mRNA
            sequence.
ACCESSION   AA573164
NID         g2347692
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 330)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
```

```
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html

          Seq primer: -40m13 fwd. ET from Amersham
          High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .330
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1071779"
                     /clone_lib="NCI_CGAP_Br2"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
     mRNA            <1. .>330
BASE COUNT       75 a     62 c     75 g    118 t
ORIGIN

  Query Match             7.0%;  Score 149;  DB 9;  Length 330;
  Best Local Similarity  83.1%;  Pred. No. 1.20e-273;
  Matches   212;  Conservative    0;  Mismatches  39;  Indels   4;  Gaps   2;

Db      78 TAAAGATAATAGGGTGGACAGAGCGGTTACCACCAGGACCAGCAGTGGGCTCAGACCACA 137
           || |||||| |||| |  || ||||| | || ||| ||| |||||| ||||||||  ||
Cp    1687 TACAGATAACAGGGCAGCAAGGGCGGTGAGCATCAGCACCGGCAGTGAGCTCAGACTGCA 1628

Db     138 GCTTGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAA-CACC---GAGACCTTC 193
           ||| ||||||||||||||||||||||||||||||||||||  ||||   |||||| ||
Cp    1627 GCTGGGAGGAGCAGCCATTGATTTTGTGGTTATGTGGCTGGAGGCACCAGCGAGACCATC 1568

Db     194 TTTTTCATAATTACCATTGGAAATACAGAGGTGTGTATTGCCCGACACATTGGATTTCAG 253
            ||| |  ||| || ||  |||| ||||||||||||  | ||||||||||||||||||
Cp    1567 CTTTCCGAAATCACTATCAGAAAGACAGAGGTGTGTGCTACCCGACACATTGGATTTCAG 1508

Db     254 CTTCTGTGCCTGTAAATTTGCACACGGTGGCAAAACATGAGTGGGAATTTCATTCTCAGA 313
           |||||| ||||| ||||| ||||| ||||| ||||| || ||||| || ||||||||||
Cp    1507 CTTCTGAGCCTGCAAATTCGCACAGGGTGGTAAAACGTGTGTGGGGATCTCATTCTCAGA 1448

Db     314 CCCTGCTGGCCCCAG 328
           |||||||||||||||
Cp    1447 CCCTGCTGGCCCCAG 1433


RESULT   10
LOCUS       AA471183       266 bp    mRNA            EST       17-JUN-1997
DEFINITION  PMY2139 KG1a Lambda Zap Express cDNA Library Homo sapiens cDNA 5',
            mRNA sequence.
ACCESSION   AA471183
NID         g2199040
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Claudio,J.O., Liew,C.C. and Hawley,R.G.
  TITLE     Hematopoietic progenitor ESTs
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Hawley RG
            Oncology Research Laboratories
            The Toronto Hospital
            CRCS-424, 67 College St., Toronto, Ontario M5G 2M1, Canada
            Tel: 416 3403834
            Fax: 416 3403453
            Email: r.hawley@utoronto.ca
            Clone was randomly picked from subtracted KG1a primary cDNA
            library.
            Seq primer: 5' GAAATTAACCCTCACTAAAGGG 3'
            High quality sequence stop: 226.
FEATURES             Location/Qualifiers
     source          1. .266
                     /organism="Homo sapiens"
                     /note="Vector: Lambda Zap Express (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; Unidirectional cloning sites:
                     EcoRI-XhoI. mRNA was purified from KG1a cell line, cDNA
                     was synthesized using an XhoI-OligodT linker primer.
                     EcoRI adaptors were ligated, followed by digestion with
                     XhoI for directional cloning into predigested Lambda Zap
                     Express."
                     /db_xref="taxon:9606"
                     /clone_lib="KG1a Lambda Zap Express cDNA Library"
                     /cell_type="promyeloblast"
                     /cell_line="KG1a"
     mRNA            <1. .>266
BASE COUNT       70 a     74 c     57 g     65 t
ORIGIN

  Query Match             6.6%;  Score 141;  DB 7;  Length 266;
  Best Local Similarity  84.2%;  Pred. No. 1.21e-254;
  Matches   223;  Conservative    0;  Mismatches  34;  Indels   8;  Gaps   8;

Db       1 TCTTAAAAATGCAATTCAAGCCTTTTGCAATGTGCTCCGATGTGACCGTGTGGCAGCCAG 60
           ||| ||||||||||||||||||||| |||||| |||| ||||||||| ||||||||||||
Qy    1312 TCTCAAAAATGCAATTCAAGCCTTTGGCAATG-GCTCAGATGTGACCATGTGGCAGCCAG 1370

Db      61 CCTTCCCAGTACAGACCACCACTGCCCACTACCCACCACTTGCCCTTCCGGGTTAAGAAC 120
           ||   ||||| |||||||||||||||| || ||  |||||| ||| |||||||| ||||||
Qy    1371 CCCCTCCAGTCCAGACCACCACTGCC-ACCACT-ACCACT-GCC-TTCCGGGTCAAGAAC 1426

Db     121 AAGCCCCTTGGGGCCAGCAGGGGTCTGAGAATGAAATTTCCACTCATGTTTTGCCACCGT 180
           ||||| || |||||||||||| |||||||||| ||  |||| || ||||| ||||| |
Qy    1427 AAGCCTCT-GGGGCCAGCAGGG-TCTGAGAATGAGATCCCCACACACGTTTTACCACCCT 1484

Db     181 GTGCAAATTTACAGGCACAGAAGCTGAAATCCAATGTGTCGGGCAT-ACACACCTCTGTA 239
           |||| ||||| ||||| ||||||||||||||||||||||||| |  ||||||||||||
Qy    1485 GTGCGAATTTGCAGGCTCAGAAGCTGAAATCCAATGTGTCGGGTAGCACACACCTCTGTC 1544

Db     240 TTTCCACTGGTAATTATGTAAAAGA 264
           ||||   | || |||  | ||| ||
Qy    1545 TTTCTGATAGTGATTTCGGAAAGGA 1569


RESULT   11
LOCUS       H12981         521 bp    mRNA            EST       27-JUN-1995
DEFINITION  yl70a10.r1 Homo sapiens cDNA clone 43207 5'.
ACCESSION   H12981
NID         g877801
KEYWORDS    EST.
SOURCE      human clone=43207 library=Soares infant brain 1NIB vector=Lafmid BA
            host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1=Not I
            Rsite2=Hind III Whole brain from a 73 days post natal female. 1st
            strand cDNA was primed with a Not I - oligo(dT) primer [5'
            AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded
            cDNA was ligated to Hind III adaptors (Pharmacia), digested with
            Not I and directionally cloned into the Not I and Hind III sites of
            the Lafmid BA vector. Library went through one round of
            normalization. Library constructed by Bento Soares and M.Fatima
            Bonaldo.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
```

```
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            GDB: G00-415-748
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 278
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .521
                      /organism="Homo sapiens"
                      /clone="43207"
BASE COUNT       95 a    154 c    144 g    119 t      9 others
ORIGIN

  Query Match             5.6%;  Score 119;  DB 18;  Length 521;
  Best Local Similarity  75.3%;  Pred. No. 6.57e-203;
  Matches   213;  Conservative    0;  Mismatches  65;  Indels   5;  Gaps   5;

Db       2 AACCATTGCCTGGGATGCTGCCAAGGCCTGCAACCTGAATGACAACTGCAAGAAGCTGCG 61
           ||| | ||||||| | || ||||||||||||||||||| | |||| ||| ||||||    |
Qy     755 AACAACTGCCTGG-ACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGAAGTACAG 813

Db      62 CTCCTCCTACATCTCCATCTGCAACCGCGAGATCTCGCCCACCGAGCGCTGCAACCGCCG 121
             ||  |||||||| ||  ||||| || | ||   | | ||| ||||  ||||||||||||
Qy     814 GTCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCAACCGCCG 870

Db     122 CAAGTGCCACAAGGCCCTGCGCCAGTTCTTCGACCGGGTGCCCAGCGAGTACACCTACCG 181
           ||||||||||||||||  | ||||||||||||  ||| ||   | || ||| |||| |
Qy     871 TAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACAGCTACGG 930

Db     182 CATGCTCTTCTGCTCCTGCCAAGACCAGGCGTGCGCTGAGCCGCGGNCAAA-ACCATCCT 240
           ||||||||||||||||||  |||   || ||| | |||| ||||  | | || ||| |
Qy     931 GATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAGCGGCGGCGACAGACTATCGT 990

Db     241 GCCCAGCTGCTCCTATGAGGACAAGGAGAAGCCCAACTGCCTG 283
           |||   |||||||||| ||    |||| ||||||||||||
       991 CCCCGTGTGCTCCTATGAAGAACGAGAGAGGCCCAACTGCCTG 1033


RESULT   12
LOCUS       R02249        229 bp    mRNA            EST       31-MAR-1995
DEFINITION  ye83h05.r1 Homo sapiens cDNA clone 124377 5'.
ACCESSION   R02249
NID         g751985
KEYWORDS    EST.
SOURCE      human clone=124377 library=Soares fetal liver spleen 1NFLS
            vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B
            (ampicillin resistant) primer=M13RP1 Rsitel=Pac I Rsite2=Eco RI
            Liver and spleen from a 20 week-post conception male fetus. 1st
            strand cDNA was primed with a Pac I - oligo(dT) primer [5'
            AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
            cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
            I and cloned into the Pac I and Eco RI sites of the modified pT7T3
            vector.  Library went through one round of normalization. Library
            constructed by Bento Soares and M.Fatima Bonaldo.
  ORGANISM  Homo sapiens
            Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 214
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES              Location/Qualifiers
     source           1. .229
                      /organism="Homo sapiens"
                      /clone="124377"
BASE COUNT       49 a     85 c     55 g     39 t      1 others
ORIGIN

  Query Match             4.7%;  Score 100;  DB 3;  Length 229;
  Best Local Similarity  75.3%;  Pred. No. 4.72e-159;
  Matches   171;  Conservative    0;  Mismatches  53;  Indels   3;  Gaps   3;

Db       1 CCAAGAGCAACCATTGCCTGGATGCTGCCAAGGCCTGCAACCTGAATGACAACTGCAAGA 60
           |||| | ||| | |||||||| || |||||||||||||||||| | |||| ||| ||||
Qy     747 CCAAAGGGAACAACTGCCTGGACGCAGCCAAGGCCTGCAACCTGGACGACACCTGTAAGA 806

Db      61 AGCTGCGCTCCTCCTACATCTCCATCTGCAACCGCGAGATCTCGCCCACCGAGCGCTGCA 120
           ||    | ||  |||||||| ||  ||||| || | ||   | | ||| ||||  |||||
Qy     807 AGTACAGGTCGGCCTACATCACCCCCTGCA-CCACCAGCA-T-GTCCAACGAGGTCTGCA 863

Db     121 ACCGCCGCAAGTGCCACAAGGCCCTGCGCCAGTTCTTCGACCGGGTGCCCAGCGAGTACA 180
           ||||||| |||||||||||||||||  | ||||||||||||  ||| ||   | || |||
Qy     864 ACCGCCGTAAGTGCCACAAGGCCCTCAGGCAGTTCTTCGACAAGGTTCCGGCCAAGCACA 923

Db     181 CCTACCGCATGCTCTTCTGCTCCTGCCAAGATCAGGCGTGCGCTGAG 227
           |||| | ||||||||||||||||||||  ||    || ||| | |||
Qy     924 GCTACGGGATGCTCTTCTGCTCCTGCCGGGACATCGCCTGCACCGAG 970


RESULT   13
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
```

```
          Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /clone="97SN1787"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match             2.9%;  Score 62;  DB 13;  Length 252;
  Best Local Similarity  12.1%;  Pred. No. 3.14e-75;
  Matches    27;  Conservative  115;  Mismatches  78;  Indels   3;  Gaps   3;

Db      18 WMTTSYBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDW 77
           :: |:::|:  :: : : ::   : :: ||:   : :|: : :: :  |    : ::|:
Cp     620 ACATGCTCCAGTAGATACGCAGACAATTCTTCTCTTTCTTCATGCCCCGCTTGCAGCGGC 561

Db      78 HYWBVBNTKVDVGNHTRCSRWRBVTR-MAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMT 136
           : ::::   :::  :|  : : :: : : :::::  |:::   :::::: ::::: ||:|
Cp     560 AGTTGTACAGAGACTTCTGCTTCAAGGCCTCCATGGCGCTACGGCACTCATCCTTGGCCT 501

Db     137 CTMWCWBHYNTKCTASGWHTSTNYDVKSSTN-TWGVTBSY-DKSMHGYWCSBBVKYHTKV 194
           | :   :   |:  |:| :    ::::  |  |:|: ::: : ::: : |:::::: |::
Cp     500 CAAGGCCGGATGTCAGGCTGAAGTTGGTTTCCTTGCCCGCCACGCACTGCCTTAGTGTGC 441

Db     195 STTRATRSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTDSKCK 237
           : |: | : | : :: ::::|:: : ::::::: |:: :::|:
Cp     440 GGTACTTGGTGCTGCAGCTCTGTTCCTTCAGGCACTGATCGCT 398


RESULT   14
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
KEYWORDS    EST.
SOURCE      rice.
  ORGANISM  Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT
            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /clone="97SN1787"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match             2.8%;  Score 60;  DB 13;  Length 252;
  Best Local Similarity   9.6%;  Pred. No. 4.79e-71;
  Matches    22;  Conservative  119;  Mismatches  86;  Indels   2;  Gaps   2;

Db      23 YBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBV 82
           :: :|  ::: : :   ::::: |: |:|:   : :  :   : :  :  ::::
Qy    1558 TTTCGGAAAGGATGGTCTCGCTGGTGCCTCCAGCCACATAACCACAAAATCAATGGCTGC 1617

Db      83 BNTKVDVGNHTRCSRWRBVTRMAH-YHDYTNCBBYNNND-YHMWHBBMYBBTGCMTCTMW 140
           :  :: :  :|:| :: :: :: : ::::   :::   : ::::::::::|||  | ::
Qy    1618 TCCTCCCAGCTGCAGTCTGAGCTCACTGCCGGTGCTGATGCTCACCGCCCTTGCTGCCCT 1677

Db     141 CWBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYWCSBBVKYHTKVSTTRAT 200
           :::: |:   :|:: :   :: ::|  : : ::::::: | :   :::::|::   :|
Qy    1678 GTTATCTGTATCGTTGGCAGAAACGTCGTAGCTGCATCCGGGAAAACAGTATGAAAAGAC 1737

Db     201 RSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTDSKCKTMWMTNKHVMTS 249
           :    ::  |::: | : :::::::::: |:: : :  |:::|  :::|:
Qy    1738 AAAAGAGAACCAAGTATTCTGTCCCTGTCCTCTTGTATATCTGAAAATC 1786


RESULT   15
LOCUS       AA270188      543 bp    mRNA            EST       26-MAR-1997
DEFINITION  vb11e06.r1 Soares mouse NML Mus musculus cDNA clone 748642 5', mRNA
            sequence.
ACCESSION   AA270188
NID         g1908517
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:457626
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
```

```
                    and Eco RI sites of the modified pT7T3 vector. Library
                    constructed and normalized by Bento Soares and M.Fatima
                    Bonaldo."
                    /db_xref="taxon:10090"
                    /clone="748642"
                    /clone_lib="Soares mouse NML"
                    /tissue_type="Liver"
                    /lab_host="DH10B"
     mRNA           <1. .>543
BASE COUNT      154 a    122 c    126 g    141 t
ORIGIN

  Query Match            2.5%;  Score 54;  DB 24;  Length 543;
  Best Local Similarity  79.3%;  Pred. No. 1.06e-58;
  Matches    73;  Conservative    0;  Mismatches  19;  Indels   0;  Gaps   0;

Db     452 GGCTGCAGTCATGTCTTTAGAATCCAACGCTATTAAATAAGATCCGGAAGAATCTCTATC 511
           |||||| | | |||||| ||||  ||| |||  | | |||||   |||||||||| ||||
Cp    2134 GGCTGCGGGCCTGTCTTCAGAAGGCAATGCTGCTGAGTAAGACATGGAAGAATCTTTATC 2075

Db     512 AATAGCATCTCCTCCTCCTAGCCAGTAGAGAG 543
           ||||| |||||||  ||||| ||||||||||||
Cp    2074 AATAGTATCTCCTGTTCCTAACCAGTAGAGAG 2043


Search completed: Wed Jun 24 23:10:00 1998
Job time : 4081 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 23 18:34:40 1998; MasPar time 15.71 Seconds
456.813 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-4
Description: (1-468) from US08866354.pep
Perfect Score: 3384
Sequence: 1 MFLATLYFALPLLDLLMSAE.....PVLMLTALAALLSVSLAETS 468

Scoring table: PAM 150
Gap 11

Searched: 124785 seqs, 15338987 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq31-2
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27

Statistics: Mean 34.847; Variance 146.333; scale 0.238

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3384	100.0	468	27	W27327	Rat glial cell derive	0.00e+00
2	95	2.8	350	15	R82686	Vascular endothelial	2.23e+01
3	96	2.8	505	1	P80885	Sequence encoded by L	1.89e+01
4	91	2.7	316	24	W15023	C-mpl ligand variant	4.31e+01
5	90	2.7	332	20	W12929	Mpl ligand analogue,	5.06e+01
6	92	2.7	369	3	P50258	Pseudorabies virus (B	3.66e+01
7	90	2.7	676	27	W35853	Human CD18 for use in	5.06e+01
8	90	2.7	769	4	R24256	Beta-subunit CD18 of	5.06e+01
9	91	2.7	947	27	W42402	Amino acid sequence o	4.31e+01
10	87	2.6	32	3	R15252	Carbohydrate binding	8.20e+01
11	88	2.6	95	18	W00601	Rat carboxypeptidase	6.99e+01
12	87	2.6	285	21	W01770	IL-3 variant/FXa/GlyS	8.20e+01
13	87	2.6	285	21	W01771	IL-3 variant/GlySer/1	8.20e+01
14	87	2.6	332	21	W01760	Variant c-mpl ligand	8.20e+01
15	87	2.6	504	3	P93707	Sequence of the gag p	8.20e+01
16	87	2.6	514	3	R15237	Fusarium oxysporum C-	8.20e+01
17	88	2.6	3211	1	P81769	Sequence encoded by t	6.99e+01
18	86	2.5	119	27	W43305	p17 antigen of human	9.61e+01
19	86	2.5	120	2	P70272	The sequence encoding	9.61e+01
20	86	2.5	132	5	R26326	HTLV-IIIb p17 sequenc	9.61e+01
21	86	2.5	169	10	R55855	HIV-1 MA delta-79-90	9.61e+01
22	86	2.5	169	10	R55851	HIV-1 MA delta-5-16 m	9.61e+01
23	86	2.5	171	10	R55857	HIV-1 MA delta-105-11	9.61e+01
24	86	2.5	267	2	P94509	Sequence encoded by p	9.61e+01
25	86	2.5	332	20	W12942	Mpl ligand analogue,	9.61e+01
26	86	2.5	363	8	R43267	p37gag encoded by vec	9.61e+01
27	86	2.5	438	13	R77183	HIV-1 gag protein p45	9.61e+01
28	86	2.5	478	13	R73878	HIV-1 gag polyprotein	9.61e+01
29	86	2.5	478	3	P60309	Sequence of gag and p	9.61e+01
30	86	2.5	500	2	P70269	The sequence encoding	9.61e+01
31	86	2.5	506	6	R29704	gag gene decoded from	9.61e+01
32	86	2.5	506	3	P60176	Sequence of ARV-2 (9B	9.61e+01
33	86	2.5	508	3	P60419	Sequence of LAV virus	9.61e+01
34	86	2.5	512	20	W02296	HIV-gag internal prot	9.61e+01
35	86	2.5	512	4	R24028	HIV-gag protein.	9.61e+01
36	86	2.5	512	2	R10866	HIV gag precursor pro	9.61e+01
37	86	2.5	515	1	P91235	(ENV-80)(GAG-VII)(Hex	9.61e+01
38	86	2.5	522	2	R04024	Gag gene product of c	9.61e+01
39	86	2.5	524	27	W43304	Modified p17 antigen	9.61e+01
40	86	2.5	600	2	P70541	HTLV-III gag/env gene	9.61e+01
41	86	2.5	607	7	R34662	Rat 5HT transporter.	9.61e+01
42	86	2.5	639	1	P91052	Sequence of fusion pr	9.61e+01
43	86	2.5	769	1	P80836	Beta subunit of human	9.61e+01
44	86	2.5	769	1	R07113	Recombinant beta-subu	9.61e+01
45	86	2.5	3077	1	P93283	Sequence of clone HIV	9.61e+01

ALIGNMENTS

RESULT 1
ID W27327 standard; Protein; 468 AA.
AC W27327;
DT 27-APR-1998 (first entry)
DE Rat glial cell derived neurotrophic factor receptor alpha.
KW Glial cell derived neurotrophic factor receptor alpha; GDNFR alpha;
KW GDNF; rat; kidney disease; glomerulonephritis; therapy.
OS Rattus sp.
FH Key Location/Qualifiers
FT Peptide 1..24
FT /label= Sig_peptide
FT Domain 25..445
FT /note= "extracellular domain"
FT Domain 446..468
FT /note= "GPI-attachment site"
FT Modified_site 61
FT /note= "O-glycosylated"
FT Modified_site 349
FT /note= "O-glycosylated"
FT Modified_site 408
FT /note= "O-glycosylated"
FT Cleavage_site 428..430
FT /note= "GPI-anchor cleavage/attachment site"
PN WO9733912-A2.
PD 18-SEP-1997.
PF 13-MAR-1997; U04363.
PR 14-MAR-1996; US-618236.
PR 14-MAR-1996; US-615902.
PA (GETH) GENENTECH INC.
PI Klein RD, Moore MW, Rosenthal A, Ryan AM;
DR WPI; 97-470819/43.
DR N-PSDB; T84975.
PT Isolated glial cell derived neurotrophic factor receptor alpha -
PT useful to develop products to diagnose and treat associated
PT disorders, particularly enteric nervous system or kidney disorders
PS Claim 1; Page 78-79; 100pp; English.
CC This protein comprises full-length rat glial cell derived
CC neurotrophic factor receptor alpha (GDNFR alpha), a novel
CC GPI-linked protein that is a ligand-binding component of the
CC receptor system for GDNF. Its amino acid sequence was deduced
CC from an isolated cDNA clone (see T84975). The invention relates
CC to novel uses of GDNF and its receptor. In particular, it
CC relates to native rat GDNFR alpha (see W27327), its variants and

```
CC   soluble derivatives (extracellular domain), chimeric GDNFR alpha
CC   and antibodies which bind to the GDNFR alpha, including agonist
CC   and neutralising antibodies, as well as various uses for these
CC   molecules.  It also relates to assay systems for detecting ligands
CC   to GDNFR alpha, systems for studying the physiological role of
CC   GDNF, diagnostic techniques for identifying GDNF-related conditions,
CC   methods for identifying molecules homologous to GDNFR alpha, and
CC   therapeutic techniques (claimed) for the treatment of GDNF-related
CC   and GDNFR alpha-related conditions, particularly kidney disease
CC   associated with glomerulonephritis and enteric nervous system
CC   related disorders.  Transgenic and knockout animals are also
CC   claimed.
SQ   Sequence   468 AA;

  Query Match          100.0%;  Score 3384;  DB 27;  Length 468;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   468;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db       1 mflatlyfalplldllmsaevsggdrldcvkasdqclkeqscstkyrtlrqcvagketnf 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 sltsgleakdecrsamealkqkslyncrckrgmkkeknclriywsmyqslqgndlledsp 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 yepvnsrlsdifravpfisdvfqqvehiskgnncldaakacnlddtckkyrsayitpctt 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db     181 smsnevcnrrkchkalrqffdkvpakhsygmlfcscrdiacterrrqtivpvcsyeerer 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db     241 pnclslqdscktnyicrsrladfftncqpesrsvsnclkenyadcllaysgligtvmtpn 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db     301 yvdssslsvapwcdcsnsgndledclkflnffkdntclknaiqafgngsdvtmwqpappv 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db     361 qtttatttfafrvknkplgpagseneipthvlppcanlqaqklksnvsgsthlclsdsdf 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     361 QTTTATTTFAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db     421 gkdglagasshittksmaappscslsslpvlmltalaallsvslaets 468
           |||||||||||||||||||||||||||||||||||||||||||||||
Qy     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    2
ID   R82686 standard; Protein; 350 AA.
AC   R82686;
DT   18-APR-1996  (first entry)
DE   Vascular endothelial growth factor 2.
KW   VEGF2; vascular endothelial growth factor; vascular permeability factor;
KW   PDGF family; angiogenesis; induce growth; bone.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1..24
FT                   /label= signal_peptide
FT   protein         25..350
FT                   /label= mature_protein
FT   misc_difference 62
FT                   /note= "conserved cys"
FT   domain          85..98
FT                   /note= "signature for the PDGF/VEGF family"
FT   misc_difference 87
FT                   /note= "conserved cys"
FT   misc_difference 93
FT                   /note= "conserved cys"
FT   misc_difference 96
FT                   /note= "conserved cys"
FT   misc_difference 97
FT                   /note= "conserved cys"
FT   misc_difference 104
FT                   /note= "conserved cys"
FT   misc_difference 140
FT                   /note= "conserved cys"
FT   misc_difference 142
FT                   /note= "conserved cys"
PN   WO9524473-A1.
PD   14-SEP-1995.
PF   12-MAY-1994; U05291.
PR   08-MAR-1994; US-207550.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Cao L,  Hu J,  Fraser CM;
DR   WPI; 95-328266/42.
DR   N-PSDB; T03950.
PT   Human vascular endothelial growth factor 2 polypeptide - used to
PT   provide diagnostic, therapeutic and/or prophylactic effects against
PT   various diseases, esp. in wound healing and periodontal disease
PS   Claim 1; Fig 1A-D; 48pp; English.
CC   This full length vascular endothelial growth factor 2 (VEGF2) comprises
CC   350 amino acids of which approx. the first 24 amino acids represent the
CC   leader sequence. VEGF2 is structurally related to the VEGF/PDGF family.
CC   It is partic. important that all eight cysteines are conserved within
CC   the members of the family.  In addition, the signature for the PDGF/VEGF
CC   family, PXCVXXXRCXGCCN, is conserved in VEGF2. VEGF2 induces growth of
CC   damaged bone, peridotium or ligament tissue.  It is used for the
CC   promotion of endothelialisation in vascular graft surgery.  Since
CC   angiogenesis is important in keeping wounds clean and non-infected, VEGF2
CC   may be used in association with surgery and following the repair of cuts.
SQ   Sequence   350 AA;

  Query Match            2.8%;  Score 95;  DB 15;  Length 350;
  Best Local Similarity  23.2%;  Pred. No. 2.23e+01;
  Matches    13;  Conservative   14;  Mismatches  26;  Indels   3;  Gaps   3;

Db       8 ywkmykcqlrkggwqhnreqanlnsrteetikfraahynteilksidnewrktqcm 63
           || ||   ]:      :     :||| : | |||  : ::::  ::: :   |:
Qy     103 YWSMYQS-LQGNDLLEDSPYEPVNSRLSD-I-FRAVPFISDVFQQVEHISKGNNCL 155


RESULT    3
ID   P80885 standard; protein; 505 AA.
AC   P80885;
DT   16-DEC-1990  (first entry)
DE   Sequence encoded by LAV MA L GAG gene
KW   HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
OS   Lymphadenopathy associated virus MA L.
PN   WO8707906-A.
PD   30-DEC-1987.
PF   22-JUN-1987; E00326.
PR   23-JUN-1986; EP-401380.
PA   (INSP) Inst Pasteur.
PI   Alizon M, Sonigo P, Wain-Hobson S, Montagnier L;
DR   WPI; 88-014396/02.
DR   N-PSDB; N80437.
PT   New variants of lymphadenopathy associated virus (LAV) -
PT   used for prodn. of DNA, antigens and antibodies used in
PT   diagnosis of AIDS and pre-AIDS
PS   Claim 8; Fig 8A-8I; 72pp; English.
CC   LAV EL I (N80436) and LAV MA L (N80437) were isolated from the peripheral
CC   blood lymphocytes of patients. Different AIDS virus isolates concerned
CC   are designated by 3 letters of the patients name. Stable probes including
CC   the DNA sequences can be used for detection of the new LAV viruses or
CC   related viruses or DNA proviruses in eg. biological samples. The proteins
CC   or peptides can be used for detection of antibodies induced in vivo and
CC   present in biological fluids. The DNA can also be used for the expression
CC   of LAV viral antigens for the prodn. of a vaccine against LAV. The
CC   polypeptides can also be used for the prodn. of antibodies for the
CC   detection of proteins related to the LAV viruses, partic. for diagnosis
CC   of AIDS or pre-AIDS.
SQ   Sequence   505 AA;
```

```
  Query Match             2.8%;  Score 96;  DB 1;  Length 505;
  Best Local Similarity  41.9%;  Pred. No. 1.89e+01;
  Matches    18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db      28 kyr-lkhlvwasrelerfalnpglletgegcqqimeql-qstl 68
           ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy      45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT    4
ID   W15023 standard; Protein; 316 AA.
AC   W15023;
DT   01-JAN-1998  (first entry)
DE   C-mpl ligand variant dimer.
KW   C-mpl ligand; thrombopoietin; receptor; agonist; cytokine; human;
KW   haematopoietic cell; stem cell; thrombocytopaenia; gene therapy;
KW   pMON32136.
OS   Chimeric Homo sapiens.
     Chimeric Synthetic.
     Key             Location/Qualifiers
     Protein         1..153
FT                   /note= "c-mpl(1-153)"
FT   Peptide         154..163
FT                   /note= "linker"
FT   Protein         164..316
FT                   /note= "c-mpl(1-153)"
PN   WO9712978-A1.
PD   10-APR-1997.
PF   04-OCT-1996; U15938.
PR   05-OCT-1995; US-004824.
PA   (SEAR ) SEARLE & CO G D.
PI   Feng Y,  Giri JG,  McKearn JP,  McWherter CA,  Pegg LE;
PI   Staten NR,  Summers NL;
DR   WPI; 97-226221/20.
DR   N-PSDB; T62980.
PT   Novel c-mpl receptor agonist polypeptide(s) - stimulate
PT   haematopoietic cell production, useful in thrombocytopenia treatment
PT   and selective ex vivo expansion of haematopoietic stem cells
PS   Example 6; Page 61-62; 151pp; English.
CC   This polypeptide is a c-mpl ligand variant comprising amino acid
CC   residues 1-153 of c-mpl ligand (see W14996) fused via a linker
CC   (W15003) to amino acids 1-153 of c-mpl ligand.  DNA construct
CC   pMON32136 (see T62980) encoding the variant can be used as a PCR
CC   template for generating novel forms of c-mpl ligand.  Specifically
CC   claimed circularly permuted variants of c-mpl ligand (see W15005-
CC   16) are useful in the treatment of thrombocytopaenia and selective
CC   ex vivo expansion of haematopoietic stem cells.
SQ   Sequence   316 AA;

  Query Match             2.7%;  Score 91;  DB 24;  Length 316;
  Best Local Similarity  22.0%;  Pred. No. 4.31e+01;
  Matches    28;  Conservative   34;  Mismatches  60;  Indels   5;  Gaps   5;

Db     152 vrefg-gnggnmaspappacdlrvlskllrdshvlhsrlsqcpevhplptpvllpavdfs 210
           :: || |:   |  |||| :    :      :| :  |:     : :|| || |  ::
Qy     342 IQAFGNGSDVTMWQPAPPV-QTTTATTT-TAFRVKNKPLGPAGSENEIPTHVLPPCANLQ 399

Db     211 lgewktqmeetkaqdilgavtlllegvmaargqlgptclssllg-qlsgqvrlllgalqs 269
             |: :  : ::  |:   :  :|: :| :::    :::  :  ||:   |:| || :
Qy     400 AQKLKSNVSGS-THLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAA 458

Db     270 llgtqlp 276
           ||:  |:
Qy     459 LLSVSLA 465


RESULT    5
ID   W12929 standard; Protein; 332 AA.
AC   W12929;
DT   16-APR-1997  (first entry)
DE   Mpl ligand analogue, [Asn54, Ser56] mpl ligand.
KW   Native human mpl ligand; altered glycosylation pattern;
KW   carbohydrate side chain; sialic acid content; thrombocytopaenia;
KW   megakaryocyte/platelet deficiency.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   modified_site   54
FT                   /label= Gln54Asn
FT   modified_site   56
FT                   /label= Glu56Ser
PN   WO9625498-A2.
PD   22-AUG-1996.
PF   13-FEB-1996; U02492.
PR   15-FEB-1995; US-388779.
PR   09-FEB-1996; US-591070.
PA   (AMGE-) AMGEN INC.
PI   Elliott SG;
DR   WPI; 96-393399/39.
PT   Mpl ligand analogue has sequence with added, deleted or altered
PT   glycosylation site(s) - useful in compsns. to treat
PT   thrombocytopaenia
PS   Claim 8; Page -; 106pp; English.
CC   The sequences given in W12921-54 represents claimed analogues of
CC   native human mpl ligand.  These analogues have altered glycosylation
CC   patterns.  Alteration of glycosylation of mpl ligand can cause
CC   improved biological activity.  The mutated proteins comprise a
CC   greater or lesser number of carbohydrate side chains and higher or
CC   lower sialic acid content than wild type mpl ligand.  Compositions
CC   comprising these mpl ligand analogues may be used to treat diseases
CC   involving an existing or expected megakaryocyte/platelet deficiency,
CC   i.e. in thrombocytopaenia.
SQ   Sequence   332 AA;

  Query Match             2.7%;  Score 90;  DB 20;  Length 332;
  Best Local Similarity  23.2%;  Pred. No. 5.06e+01;
  Matches    22;  Conservative   29;  Mismatches  42;  Indels   2;  Gaps   2;

Db      20 hvlhsrlsqcpevhplptpvllpavdfslgewktnmsetkaqdilgavtlllegvmaarg 79
           :| :  |:     : :|| || |  ::      |:|:| : ::  |:   :  :|: :| :
Qy     372 RVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCLSDSDFGKDGLAGASS 430

Db      80 qlgptclssllg-qlsgqvrlllgalqsllgtqlp 113
           ::    :::  :  ||:   |:| || :||:  |:
Qy     431 HITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLA 465


RESULT    6
ID   P50258 standard; Protein; 369 AA.
AC   P50258;
DT   22-NOV-1991  (first entry)
DE   Pseudorabies virus (BUK-7) thymidine kinase.
KW   Pseudorabies virus (BUK-7); thymidine kinase; mutagenesis;
KW   temp. resistance; vaccine.
OS   Pseudorabies virus.
PN   US4514497-A.
PD   30-APR-1985.
PF   30-DEC-1983; 567018.
PR   30-DEC-1983; US-567018.
PA   (BAYU) Baylor College of Medicine.
PA   (NOVA-) Novagene Ltd.
PI   Kit M, Kit S;
DR   WPI; 85-122203/20.
DR   N-PSDB; N50286.
PT   Temp.-resistant pseudorabies virus - useful in vaccines as virus
PT   does not produce thymidine kinase.
PS   Disclosure; Fig. 5; 26pp; English.
CC   The pseudorabies virus (BUK-7)CC thymidine kinase coding sequence can
CC   be mutated using a mutagen or deleted, so producing a virus
CC   (which is also modified to be temperature sensitive) which cannot
CC   produce functional thymidine kinase. Such a mutant is useful in a
CC   vaccine to control the spread of pseudorabies virus. Vaccinated
CC   animals are also less likely to become carriers of the virus, and are
CC   unlikely to require a dormant infection with pathogenic field
CC   strains. The mutagen-induced mutant is pseudorabies virus (BUK 5A0
CC   ATCC VR-2078, and may be lyophilised for storage.
```

```
SQ   Sequence   369 AA;

  Query Match            2.7%;  Score 92;  DB 3;  Length 369;
  Best Local Similarity  23.9%;  Pred. No. 3.66e+01;
  Matches    11;  Conservative   15;  Mismatches  19;  Indels   1;  Gaps   1;

Db    255 rnvyamlvntsrylssgrrwrddwgraprfdqttr-dclalnelcr 299
          :: |:||  : | ::   |:|:       :::  | :||:| : |:
Qy    206 KHSYGMLFCSCRDIACTERRRQTIVPVCSYEERERPNCLSLQDSCK 251


RESULT    7
ID   W35853 standard; protein; 676 AA.
AC   W35853;
DT   27-APR-1998  (first entry)
DE   Human CD18 for use in T lymphocyte veto molecule.
KW   Human; CD18; T lymphocyte veto molecule; chimeric molecule;
KW   targetting polypeptide; suppression; immune response; treatment;
KW   autoimmune disease; allergy; immunological disorder;
KW   transplant rejection.
OS   Homo sapiens.
PN   WO9737687-A1.
PD   16-OCT-1997.
PF   10-APR-1997; U05943.
PR   10-APR-1996; US-630172.
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
PI   Staerz UD;
DR   WPI; 97-512419/47.
PT   T lymphocyte veto molecule comprising response cell activating
PT   protein - linked to molecule that targets stimulator cell marker,
PT   used for selective suppression of immune response, e.g. prevention
PT   of graft rejection or treatment of auto-immune disease
PS   Claim 37; Pages 65-68; 309pp; English.
CC   A novel T lymphocyte veto molecule is a chimeric molecule
CC   comprising a protein, e.g. the present sequence, linked to a
CC   targetting polypeptide that binds a molecule, which differentiates
CC   a host cell from a tissue graft cell, or selectively targets a
CC   stimulator cell involved in the autoimmune response.
CC   A veto molecule, in which the protein binds a molecule that targets
CC   stimulator cells, can be used to suppress an immune response and
CC   therefore treat autoimmune diseases, e.g. systemic lupus
CC   erythematosus, myasthenia gravis, rheumatoid arthritis, insulin
CC   dependent diabetes mellitus, multiple sclerosis, coeliac disease,
CC   autoimmune thyroiditis, Addison's or Grave's diseases and
CC   rheumatoid carditis, allergies and other immunological disorders.
CC   Where the protein binds a molecule that differentiates graft and
CC   host cells, the veto molecule can be used to reduce transplant
CC   rejection. The veto molecule provides specific regulation of
CC   particular stimulator cells that can kill graft cells or respond
CC   to autoantigens, but leave other stimulator cells unaffected, e.g.
CC   CD4 or CD8 positive cells can be regulated without one affecting
CC   the other. The veto molecule can be administered locally to
CC   minimise generalised immunosuppression.
SQ   Sequence   676 AA;

  Query Match            2.7%;  Score 90;  DB 27;  Length 676;
  Best Local Similarity  29.3%;  Pred. No. 5.06e+01;
  Matches    24;  Conservative   22;  Mismatches  24;  Indels  12;  Gaps   8;

Db    451 dtqyigkncecqtqgrssqelegscrkdnns-ii-cs-glgdcvcgqclchts-----d- 501
          : : |:|  :|   :::  :|: :|:|  :: |  |: :::: || :  || :     |
Qy    144 QVEHISKGNNCLDAAKAC-NLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDK 202

Db    502 vpgkliygq-ycecdtinc-er 521
          ||:|  ||  :| |  | | ||
Qy    203 VPAKHSYGMLFCSCRDIACTER 224


RESULT    8
ID   R24256 standard; Protein; 769 AA.
AC   R24256;
DT   15-JUL-1992  (first entry)
DE   Beta-subunit CD18 of leukocyte integrin.
KW   Cell adhesion; intercellular adhesion molecule; endothelium;
KW   inflammation; ICAM-1; leukocyte chemotaxis; rhinovirus.
FH   Key             Location/Qualifiers
FT   peptide         1..22
FT                   /label= signal
FT   protein         23..769
FT   region          125..150
FT                   /note= "Peptide 1 (see R22104) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
FT   region          165..189
FT                   /note= "Peptide 6 (see R22109) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
FT   region          221..244
FT                   /note= "Peptide 5 (see R22108) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
FT   region          250..278
FT                   /note= "Peptide 4 (see R22107) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
FT   region          299..323
FT                   /note= "Peptide 2 (see R22105) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
FT   region          337..365
FT                   /note= "Peptide 3 (see R22106) tested for activity in
FT                   leukocyte/endothelial cell adhesion assay"
PN   WO9203473-A.
PD   05-MAR-1992.
PF   23-AUG-1991; UO6063.
PR   27-AUG-1990; US-573624.
PA   (CETU ) CETUS CORP.
PI   Liu DY, Kaymakcalan Z, Mundy K;
DR   WPI; 92-096833/12.
DR   N-PSDB; Q22780.
PT   Peptide(s) derived from beta sub-unit CD18 of leukocyte integrins
PT   - prevent leukocyte binding to ICAM and leukocyte chemotaxis, for
PT   treating inflammatory diseases and rhinoviral infection
PS   Disclosure; Fig 1; 31pp; English.
CC   This sequence is the beta-subunit, CD18, of the leukocyte
CC   integrins. Peptides were designed based on the beta-subunit sequence
CC   and were tested for activity in the leukocyte/endothelial cell
CC   adhesion assay. See R22104-R22112.
SQ   Sequence   769 AA;

  Query Match            2.7%;  Score 90;  DB 4;  Length 769;
  Best Local Similarity  30.8%;  Pred. No. 5.06e+01;
  Matches    24;  Conservative   20;  Mismatches  22;  Indels  12;  Gaps   8;

Db    477 igkncecqtqgrssqelegscrkdnns-ii-cs-glgdcvcgqslchtt-----d-vpgk 527
          |:|  :|   :::  :|: :|:|  :: |  |: :::: || :  || :     | ||:|
Qy    148 ISKGNNCLDAAKAC-NLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAK 206

Db    528 liygq-ycecdtinc-er 543
            ||  :| |  | | ||
Qy    207 HSYGMLFCSCRDIACTER 224


RESULT    9
ID   W42402 standard; Protein; 947 AA.
AC   W42402;
DT   15-APR-1998  (first entry)
DE   Amino acid sequence of NF-kappaB inducing kinase.
KW   Human tumour necrosis factor receptor-associated factor 2; TRAF2;
KW   TRAF-2 binding protein; NF-kappaB activity; NF-kappaB induction;
KW   intracellular signalling activity; acute hepatitis;
KW   autoimmune-induced cell death.
OS   Homo sapiens.
PN   WO9737016-A1.
PD   09-OCT-1997.
PF   01-APR-1997; IL0117.
PR   26-AUG-1996; IL-119133.
PR   02-APR-1996; IL-117800.
PA   (YEDA ) YEDA RES & DEV CO LTD.
PI   Boldin M, Kovalenko A, Malinin N, Mett I, Wallach D;
DR   WPI; 97-503101/46.
```

```
DR   N-PSDB; V03325.
PT   DNA encoding tumour necrosis factor receptor-associated factor
PT   binding molecule - used for modulation or mediation in cells of the
PT   activity of NF-kB
PS   Disclosure; Fig 6; 127pp; English.
CC   The present sequence represents a NF-kappa inducing kinase (NIK). The
CC   full length DNA sequence was obtained by PCR using clone 10 (V03324).
CC   NIK was found to induce NF-kappaB even more effectively than TRAF2.
CC   Proteins capable of binding to TRAF2 and NIK were identified.
CC   The NIK or TRAF-2 binding proteins can be used for modulation or
CC   mediation in cells of NF-kappaB activity or any other intracellular
CC   signalling activity modulated or mediated by TRAF2. TRAF-binding proteins
CC   are especially used for prevention or treatment of pathological
CC   conditions associated with NF-kappaB induction, e.g. acute hepatitis,
CC   autoimmune-induced cell death, e.g. death of the beta Langerhans cells or
CC   the pancreas that results in diabetes, the death of cells in graft
CC   rejection, the death of oligodendrocytes in the brain in multiple
CC   sclerosis, and AIDS-inhibited T cell suicide which causes proliferation
CC   of the AIDS virus and hence the AIDS disease. The proteins are also
     useful for screening of ligands capable of binding to a protein, which
     are useful for modulating cellular activity modulated/mediated by TRAF2.
     Sequence  947 AA;

  Query Match             2.7%;  Score 91;  DB 27;  Length 947;
  Best Local Similarity  28.8%;  Pred. No. 4.31e+01;
  Matches   21;  Conservative   10;  Mismatches  37;  Indels   5;  Gaps   4;

Db    540 clqpdglgkslltgdyipgtethmapevvlgrscdakvdvwssccmmlhmlngchpwtqf 599
          ||  :  :  ||:  |     | |:|: |   |    |  |  |    : |: |     |
Qy    277 CLK-ENYADCLLA--YSGLIGTVMTPNYV-DSS-SLSVAPWCDCSNSGNDLEDCLKFLNF 331

Db    600 frgplclkiasep 612
          |:   ||| | ::
Qy    332 FKDNTCLKNAIQA 344


RESULT   10
ID   R15252 standard; Protein; 32 AA.
AC   R15252;
DT   12-FEB-1992  (first entry)
DE   Carbohydrate binding domain #9.
KW   cellulose; CBD; hemicellulosic substrate;
KW   Trichoderma reesei; cellulase; terminal A region.
PN   WO9117244-A.
PD   14-NOV-1991.
PF   08-MAY-1991; DK0124.
PR   09-MAY-1990; DK-001158.
PA   (NOVO ) NOVO NORDISK A/S.
PI   Woldike HF, Hagen F, Hjort CM, Hastrup S;
DR   WPI; 91-353766/48.
PT   New fungal (hemi)cellulose degrading enzymes - for prodn. of liq.
PT   fuel gas and feed protein, have specified carbohydrate binding domain
PS   Claim 20; Page 45; 73pp; English.
CC   This CBD is homologous to a terminal A region of Trichoderma reesei
CC   cellulases and effects binding of a protein to an insoluble
CC   cellulosic or hemicellulosic substrate. It is one of ten specific
CC   CBD's (see R15244-R15253) which correspond to the generic CBD
CC   formulae in R15242 and R15243. The CBD is incorporated into a fusion
CC   protein comprising a catalytic domain from a cellulase, e.g. a
CC   Bacillus endoglucanase, and optionally comprising a linking B domain
CC   from e.g. a fungal endoglucanase.
SQ   Sequence  32 AA;

  Query Match             2.6%;  Score 87;  DB 3;  Length 32;
  Best Local Similarity  35.0%;  Pred. No. 8.20e+01;
  Matches    7;  Conservative    8;  Mismatches   4;  Indels   1;  Gaps   1;

Db     14 tckspftck-kindfysqcq 32
          :||: : |: :: ||:: ||
Qy    249 SCKTNYICRSRLADFFTNCQ 268


RESULT   11
ID   W00601 standard; Protein; 95 AA.
AC   W00601;
DT   09-NOV-1996  (first entry)
DE   Rat carboxypeptidase B activation peptide.
KW   Carboxypeptidase-B; pro-enzyme; protease; insulin.
OS   Rattus sp.
PN   WO9623064-A1.
PD   01-AUG-1996.
PF   25-JAN-1996; U00995.
PR   25-JAN-1995; US-378233.
PA   (BIOT-) BIOTECHNOLOGY GEN CORP.
PI   Fulga N,  Gorecki M,  Hartman J,  Mendelovitch S;
DR   WPI; 96-362688/36.
DR   N-PSDB; T35759.
PT   Purified active recombinant carboxypeptidase B prodn. - by
PT   expressing DNA encoding the pro-enzyme, folding and enzymatic
PT   cleavage to give active form, used e.g. for insulin prodn.
PS   Example 1; Page 36; 49pp; English.
CC   The amino acid sequence of the activation peptide (W00601) of
CC   rat pro-carboxypeptidase B (pro-CPB) was deduced from a cDNA
CC   sequence (see also T35759) obtd. by PCR amplification of Sprague-
CC   Dawley rat pancreas cDNA.  The sequence of the mature enzyme (W00602)
CC   has also been deduced.  Pro-CPB can be expressed in host cells, e.g.
CC   Escherichia coli, and subsequently recovered, refolded and cleaved
CC   with trypsin to yield the active enzyme.  CPB produced this way is
CC   cheaper than porcine pancreatic enzyme, and is free of other
CC   proteases.
SQ   Sequence  95 AA;

  Query Match             2.6%;  Score 88;  DB 18;  Length 95;
  Best Local Similarity  34.3%;  Pred. No. 6.99e+01;
  Matches   12;  Conservative    8;  Mismatches  14;  Indels   1;  Gaps   1;

Db     24 nliqelantkeidfwkpdsatqvk-plttvdfhvk 57
          | || ::|  ::  | | :: |   : ||  |:||
Qy    340 NAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVK 374


RESULT   12
ID   W01770 standard; Protein; 285 AA.
AC   W01770;
DT   12-MAY-1997  (first entry)
DE   IL-3 variant/FXa/GlySer/1-153 c-mpl ligand fusion protein.
KW   c-mpl ligand; Megakaryocyte Growth and Development Factor; MGDF;
KW   thromobopoietin; TPO; variant; mutein; chimera; increased activity;
KW   decreased side effect; ex vivo expansion; stem cell; treatment;
KW   haematopoietic disorder; gene therapy; human; interleukin-3.
OS   Synthetic.
PN   WO9623888-A1.
PD   08-AUG-1996.
PF   01-FEB-1996; U00830.
PR   03-FEB-1995; US-383035.
PA   (SEAR ) SEARLE & CO G D.
PI   Baum CM,  Favara JP,  Kahn LE,  Mckearn JP,  Pegg LE;
PI   Staten NR;
DR   WPI; 96-371436/37.
DR   N-PSDB; T59355.
PT   Mutant c-mpl ligands - used for stimulating the prodn. of
PT   haematopoietic cells and in the treatment of haematopoietic
PT   disorders
PS   Example 12; Page -; 74pp; English.
CC   This protein comprises interleukin-3 variant 13288 fused to FXa, a
CC   synthetic linker and amino acids 1-153 of c-mpl ligand. It is encoded
CC   by E. coli expression plasmid, pMON26460. The native c-mpl ligand is
CC   also referred to as Megakaryocyte Growth and Development Factor (MGDF)
CC   or thromobopoietin (TPO). Variants and chimera of c-mpl ligand can
CC   have an improved biological profile, such as increased activity and/or
CC   decreased side effects, and/or improved physical properties, such as
CC   improved half-life, stability and/or re-fold efficiencies. They can be
CC   used for selective ex vivo expansion of stem cells, for the treatment of
CC   patients having a haematopoietic disorder or in human gene therapy.
SQ   Sequence  285 AA;
```

```
  Query Match           2.6%;  Score 87;  DB 21;  Length 285;
  Best Local Similarity  22.0%;  Pred. No. 8.20e+01;
  Matches   27;  Conservative   33;  Mismatches  59;  Indels   4;  Gaps   4;

Db    124 gsggggsnmaspappacdlrvlskllrdshvlhsrlsqcpevhplptpvllpavdfslgew 183
          |:|:   |  ||||  :    :      :| :  |:     : :|| || |  ::
Qy    346 GNGSDVTMWQPAPPV-QTTTATTT-TAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKL 403

Db    184 ktqmeetkaqdilgavtlllegvmaargqlgptclssllg-qlsgqvrlllgalqslgt 242
          |: :  : ::  |:   :  :|: :| :::    :::  :  ||:   |:| || :||:
Qy    404 KSNVSGS-THLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSV 462

Db    243 qlp 245
          |:
Qy    463 SLA 465


RESULT   13
     W01771 standard; Protein; 285 AA.
     W01771;
     12-MAY-1997  (first entry)
DE   IL-3 variant/GlySer/1-153 c-mpl ligand fusion protein.
KW   c-mpl ligand; Megakaryocyte Growth and Development Factor; MGDF;
KW   thromobopoietin; TPO; variant; mutein; chimera; increased activity;
KW   decreased side effect; ex vivo expansion; stem cell; treatment;
KW   haematopoietic disorder; gene therapy; human; interleukin-3.
OS   Synthetic.
PN   WO9623888-A1.
PD   08-AUG-1996.
PF   01-FEB-1996; U00830.
PR   03-FEB-1995; US-383035.
PA   (SEAR ) SEARLE & CO G D.
PI   Baum CM,  Favara JP,  Kahn LE,  Mckearn JP,  Pegg LE;
PI   Staten NR;
DR   WPI; 96-371436/37.
DR   N-PSDB; T59356.
PT   Mutant c-mpl ligands - used for stimulating the prodn. of
PT   haematopoietic cells and in the treatment of haematopoietic
PT   disorders
PS   Example 13; Page -; 74pp; English.
CC   This protein comprises interleukin-3 variant 13288 fused to a
CC   synthetic linker and amino acids 1-153 of c-mpl ligand. It is encoded
CC   by E. coli expression plasmid, pMON26471. The native c-mpl ligand is
CC   also referred to as Megakaryocyte Growth and Development Factor (MGDF)
CC   or thromobopoietin (TPO). Variants and chimera of c-mpl ligand can
CC   have an improved biological profile, such as increased activity and/or
CC   decreased side effects, and/or improved physical properties, such as
CC   improved half-life, stability and/or re-fold efficiencies. They can be
CC   used for selective ex vivo expansion of stem cells, for the treatment of
CC   patients having a haematopoietic disorder or in human gene therapy.
SQ   Sequence   285 AA;

  Query Match           2.6%;  Score 87;  DB 21;  Length 285;
  Best Local Similarity  22.0%;  Pred. No. 8.20e+01;
  Matches   27;  Conservative   33;  Mismatches  59;  Indels   4;  Gaps   4;

Db    124 gsggggsnmaspappacdlrvlskllrdshvlhsrlsqcpevhplptpvllpavdfslgew 183
          |:|:   |  ||||  :    :      :| :  |:     : :|| || |  ::
Qy    346 GNGSDVTMWQPAPPV-QTTTATTT-TAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKL 403

Db    184 ktqmeetkaqdilgavtlllegvmaargqlgptclssllg-qlsgqvrlllgalqslgt 242
          |: :  : ::  |:   :  :|: :| :::    :::  :  ||:   |:| || :||:
Qy    404 KSNVSGS-THLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSV 462

Db    243 qlp 245
          |:
Qy    463 SLA 465


RESULT   14
ID   W01760 standard; protein; 332 AA.
AC   W01760;
DT   29-APR-1997  (first entry)
DE   Variant c-mpl ligand (1-332) P46L+W200R.
KW   c-mpl ligand; Megakaryocyte Growth and Development Factor; MGDF;
KW   thromobopoietin; TPO; variant; mutein; chimera; increased activity;
KW   decreased side effect; ex vivo expansion; stem cell; treatment;
KW   haematopoietic disorder; gene therapy; human.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_difference 281..300
FT                   /note= "encoded by T59343, missing from the protein
FT                   sequence referred to in the specification"
FT   misc_difference 46
FT                   /note= "substitution of Leucine for wild type residue"
PN   WO9623888-A1.
PD   08-AUG-1996.
PF   01-FEB-1996; U00830.
PR   03-FEB-1995; US-383035.
PA   (SEAR ) SEARLE & CO G D.
PI   Baum CM,  Favara JP,  Kahn LE,  Mckearn JP,  Pegg LE;
PI   Staten NR;
DR   WPI; 96-371436/37.
DR   N-PSDB; T59343.
PT   Mutant c-mpl ligands - used for stimulating the prodn. of
PT   haematopoietic cells and in the treatment of haematopoietic
PT   disorders
PS   Example 2; Page -; 74pp; English.
CC   This protein comprises amino acids 1-332 of c-mpl ligand with a P46L
CC   (sic) +W200R substitution. It is encoded by BHK expression plasmid
CC   pMON26451-4 (T59343). It is a specific example of the variant c-mpl
CC   ligands of the invention. The native c-mpl ligand is also referred to
CC   as Megakaryocyte Growth and Development Factor (MGDF) or thromobopoietin
CC   (TPO). Variants and chimera of c-mpl ligand can have an improved
CC   biological profile, such as increased activity and/or decreased side
CC   effects, and/or improved physical properties, such as improved half-life,
CC   stability and/or re-fold efficiencies. They can be used for selective ex
CC   vivo expansion of stem cells, for the treatment of patients having a
CC   haematopoietic disorder or in human gene therapy.
SQ   Sequence   332 AA;

  Query Match           2.6%;  Score 87;  DB 21;  Length 332;
  Best Local Similarity  22.1%;  Pred. No. 8.20e+01;
  Matches   21;  Conservative   28;  Mismatches  44;  Indels   2;  Gaps   2;

Db     20 hvlhsrlsqcpevhplptpvllpavdlslgewktqmeetkaqdilgavtlllegvmaarg 79
          :| :  |:     : :|| || |  :|     |: :  : ::  |:   :  :|: :| :
Qy    372 RVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCLSDSDFGKDGLAGASS 430

Db     80 qlgptclssllg-qlsgqvrlllgalqslgtqlp 113
          ::   :::  :  ||:   |:| || :||:  |:
Qy    431 HITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLA 465


RESULT   15
ID   P93707 standard; Protein; 504 AA.
AC   P93707;
DT   11-MAR-1992  (first entry)
DE   Sequence of the gag protein precursor.
KW   Vaccine; diagnosis; AIDS; LAV; HTLV-III.
OS   Human immunodeficiency virus.
PN   EP-345242-A.
PD   06-DEC-1989.
PF   02-JUN-1989; 870082.
PR   03-JUN-1988; US-202271.
PR   30-JUN-1989; AU-037256.
PR   11-JUL-1989; WO-U02415.
PA   (SMIK ) SMITHKLINE BIOLOGIC.
PI   Gheysen DA, Jacobs E;
DR   WPI; 89-358596/49.
DR   N-PSDB; N92575.
PT   New recombinant DNA encoding HIV gag precursor protein - without
PT   flanking sequences, expressed in insect, yeast and mammalian
PT   cells, producing particles useful in vaccines and diagnosis
PS   Disclosure; Page 4-6; 17pp; French.
CC   The inventors claim recombinant gag protein precursor (P93707) and
```

```
CC   the DNA encoding it (N92575) operatively linked to regulatory
CC   elements functional in eukaryotic cells.
SQ   Sequence   504 AA;

  Query Match             2.6%;  Score 87;  DB 3;  Length 504;
  Best Local Similarity  37.2%;  Pred. No. 8.20e+01;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   5;  Gaps   5;

Db      28 kyk-lkhivwasrelerfavnpglletsegcreilgql-qpsl 68
           ||: |:: | |::| : |:: :|| |: : ||  :  | | ||
Qy      45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84

Search completed: Tue Jun 23 18:36:58 1998
Job time : 138 secs.
```

THIS PAGE BLANK (USPTO)

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[illegible] on: Tue Jun 23 18:32:21 1998; MasPar time 23.69 Seconds
721.562 Million cell updates/sec
[illegible]bular output not generated.

Title: >US-08-866-354-4
Description: (1-468) from US08866354.pep
Perfect Score: 3384
Sequence: 1 MFLATLYFALPLLDLLMSAE.....PVLMLTALAALLSVSLAETS 468

Scoring table: PAM 150
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir56
1:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 46.167; Variance 82.676; scale 0.558

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	178	5.3	24	2	S69080	glycosyl-phosphatidyl	4.46e-14
2	108	3.2	397	2	S70987	dnaN protein - Mycoba	2.23e-02
3	102	3.0	249	2	S09868	hypothetical protein	1.60e-01
4	103	3.0	399	2	S70983	dnaN protein - Mycoba	1.16e-01
5	100	3.0	522	2	S41819	nucleoporin p62 - hum	3.03e-01
6	99	2.9	352	2	S60024	bradykinin B1 recepto	4.17e-01
7	95	2.8	169	2	I64089	protein-export protei	1.44e+00
8	96	2.8	769	1	JC1121	leukocyte adhesion pr	1.06e+00
9	96	2.8	826	2	A60385	monocyte surface anti	1.06e+00
10	94	2.8	857	2	S33821	median body protein -	1.96e+00
11	94	2.8	1186	2	S72229	meiotic recombination	1.96e+00
12	92	2.7	100	2	S26728	hypothetical protein	3.57e+00
13	91	2.7	316	2	S58719	probable membrane pro	4.80e+00
14	90	2.7	327	2	A55356	urokinase-type plasmi	6.45e+00
15	91	2.7	336	2	D70030	alkanal monooxygenase	4.80e+00
16	92	2.7	346	2	A70144	hypothetical protein	3.57e+00
17	91	2.7	469	2	I37451	HBF-G2 (HFK-2) protei	4.80e+00
18	90	2.7	475	2	S49886	probable membrane pro	6.45e+00
19	90	2.7	493	2	JC5621	epidermal growth fact	6.45e+00
20	93	2.7	600	2	S07638	spore coat protein SP	2.65e+00
21	90	2.7	603	2	S75664	hypothetical protein	6.45e+00
22	92	2.7	688	1	C1HUS	complement subcompone	3.57e+00
23	93	2.7	695	2	S05008	serine proteinase, ca	2.65e+00
24	91	2.7	772	2	S32659	integrin beta 2 chain	4.80e+00
25	93	2.7	798	2	S09867	hypothetical protein	2.65e+00
26	90	2.7	1075	2	S54067	probable membrane pro	6.45e+00
27	87	2.6	293	2	B26637	neurogenic repetitive	1.53e+01
28	88	2.6	378	2	S00842	leukosialin precursor	1.15e+01
29	87	2.6	411	2	H69158	LPS biosynthesis RfbU	1.53e+01
30	88	2.6	415	2	A32129	carboxypeptidase B (E	1.15e+01
31	88	2.6	448	2	S57909	probable histidine pr	1.15e+01
32	87	2.6	477	1	B46396	transcription factor	1.53e+01
33	89	2.6	491	2	S49779	CDC1 protein - yeast	8.63e+00
34	88	2.6	506	1	A38068	gag polyprotein - hum	1.15e+01
35	87	2.6	525	2	A35596	nuclear pore glycopro	1.53e+01
36	89	2.6	563	1	VCMVM7	env polyprotein - bab	8.63e+00
37	88	2.6	743	2	S48917	probable regulatory p	1.15e+01
38	89	2.6	770	2	S04847	leukocyte adhesion pr	8.63e+00
39	89	2.6	771	2	A45839	leukocyte adhesion pr	8.63e+00
40	88	2.6	1188	2	F64367	pyruvate,water dikina	1.15e+01
41	87	2.6	2139	2	A35672	crumbs protein - frui	1.53e+01
42	88	2.6	2318	2	S45306	notch 3 protein - mou	1.15e+01
43	89	2.6	2524	2	A35844	Xotch protein - Afric	8.63e+00
44	86	2.5	120	5	1TAM	hiv-1 matrix protein	2.03e+01
45	86	2.5	494	2	G64382	acetolactate synthase	2.03e+01

ALIGNMENTS

RESULT 1
ENTRY S69080 #type complete
TITLE glycosyl-phosphatidylinositol-linked protein GDNFR-alpha - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 13-Mar-1998
ACCESSIONS S69080
REFERENCE S69080
#authors Treanor, J.J.S.; Goodman, L.; de Sauvage, F.; Stone, D.M.; Poulsen, K.T.; Beck, C.D.; Gray, C.; Armanini, M.P.; Pollock, R.A.; Hefti, F.; Phillips, H.S.; Goddard, A.; Moore, M.W.; Buj-Bello, A.; Davies, A.M.; Asai, N.; Takahashi, M.; Vandlen, R.; Henderson, C.E.; Rosenthal, A.
#journal Nature (1996) 382:80-83
#title Characterization of a multicomponent receptor for GDNF.
#accession S69080
##status preliminary
##molecule_type mRNA
##residues 1-24 ##label TRE
SUMMARY #length 24 #molecular-weight 2573 #checksum 2584

Query Match 5.3%; Score 178; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 4.46e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MFLATLYFALPLLDLLMSAEVSGG 24
||||||||||||||||||||||||
Qy 1 MFLATLYFALPLLDLLMSAEVSGG 24

RESULT 2
ENTRY S70987 #type complete
TITLE dnaN protein - Mycobacterium smegmatis
ORGANISM #formal_name Mycobacterium smegmatis
DATE 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 20-Mar-1998
ACCESSIONS S70987
REFERENCE S70980
#authors Salazar, L.; Fsihi, H.; de Rossi, E.; Riccardi, G.; Rios, C.; Cole, S.T.; Takiff, H.E.
#journal Mol. Microbiol. (1996) 20:283-293
#title Organization of the origins of replication of the chromosomes of Mycobacterium smegmatis, Mycobacterium leprae and Mycobacterium tuberculosis and isolation of a functional origin from M. smegmatis.
#accession S70987

```
     ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
     ##molecule_type DNA
     ##residues      1-397 ##label SAL
     ##cross-references EMBL:X92503
     ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, October 1995
GENETICS
   #gene         dnaN
   #start_codon  GTG
CLASSIFICATION   #superfamily DNA-directed DNA polymerase III beta chain
SUMMARY          #length 397  #molecular-weight 41324  #checksum 6687

  Query Match             3.2%;  Score 108;  DB 2;  Length 397;
  Best Local Similarity  25.0%;  Pred. No. 2.23e-02;
  Matches    27;  Conservative   27;  Mismatches  50;  Indels   4;  Gaps   4;

Db    172 ESVVLAATDRFRLAVRELTWVTTAGDVEAAVLVPAKTL-AEAAKAGTDGNQVHLALGSGA 230
          ::   ::|  ||:  : |    :  :: : || |   | |:  |:   |:  || | | :
Qy    361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCL-SDS 418

Db    231 SVGKDGLLGIRSE-GKRSTTRLLDAEFPKFRQLLPAEHTAVATIGVAE 277
            ||||| |  |    :| :      :: :  |: :  :|: ::::||
Qy    419 DFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAE 466


RESULT    3
ENTRY            S09868     #type complete
TITLE            hypothetical protein UL103 - human cytomegalovirus (strain
                   AD169)
ORGANISM         #formal_name human cytomegalovirus, human herpesvirus 5
   #note         host Homo sapiens (man)
DATE             07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
                   09-Sep-1997
ACCESSIONS       S09868
REFERENCE        S09749
   #authors      Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.;
                   Cerny, R.; Horsnell, T.; Hutchison III, C.A.; Kouzarides,
                   T.; Martignetti, J.A.; Preddie, E.; Satchwell, S.C.;
                   Tomlinson, P.; Weston, K.M.; Barrell, B.G.
   #journal      Curr. Top. Microbiol. Immunol. (1990) 154:125-169
   #title        Analysis of the protein-coding content of the sequence of
                   human cytomegalovirus strain AD169.
   #cross-references MUID:90269039
   #accession    S09868
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-249 ##label CHE
      ##cross-references EMBL:X17403; NID:g59591; PID:g1780882
      ##note          this sequence was submitted to the EMBL Data Library,
                       December 1989
SUMMARY          #length 249  #molecular-weight 28636  #checksum 1456

  Query Match             3.0%;  Score 102;  DB 2;  Length 249;
  Best Local Similarity  38.3%;  Pred. No. 1.60e-01;
  Matches    18;  Conservative    7;  Mismatches  20;  Indels   2;  Gaps   2;

Db    106 CTVISCVENCNLTRKCLHDLLQYLDAVNVRESFGRLLHHSARRLICS 152
          ||     | ||  |||   | |::| |  : |:| :|  | | : |:
Qy    178 CTTSMSNEVCN-RRKCHKALRQFFDKVPAKHSYG-MLFCSCRDIACT 222


RESULT    4
ENTRY            S70983     #type complete
TITLE            dnaN protein - Mycobacterium tuberculosis
ORGANISM         #formal_name Mycobacterium tuberculosis
DATE             12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
                   20-Mar-1998
ACCESSIONS       S70983
REFERENCE        S70980
   #authors      Salazar, L.; Fsihi, H.; de Rossi, E.; Riccardi, G.; Rios, C.;
                   Cole, S.T.; Takiff, H.E.
   #journal      Mol. Microbiol. (1996) 20:283-293
   #title        Organization of the origins of replication of the chromosomes
                   of Mycobacterium smegmatis, Mycobacterium leprae and
                   Mycobacterium tuberculosis and isolation of a functional
                   origin from M. smegmatis.
   #accession    S70983
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-399 ##label SAL
      ##cross-references EMBL:X92504
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, October 1995
GENETICS
   #gene         dnaN
CLASSIFICATION   #superfamily DNA-directed DNA polymerase III beta chain
SUMMARY          #length 399  #molecular-weight 41773  #checksum 9406

  Query Match             3.0%;  Score 103;  DB 2;  Length 399;
  Best Local Similarity  27.7%;  Pred. No. 1.16e-01;
  Matches    28;  Conservative   27;  Mismatches  42;  Indels   4;  Gaps   4;

Db    178 ATDSRLAVRELEWSASSPDIEAAVLVPAKTL-AEAAKAGIGGSDVRLSLGTGPGVGKDGL 236
          :|  |:  : |  ::|  :| : || |   | |:  |: ::||  :| | : :  |||||
Qy    368 TTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCL-SDSDFGKDGL 425

Db    237 LGISGN-GKRSTTRLLDAEFPKFRQLLPTEHTAVATMDVAE 276
            | |::   :| :      :: :  |: |  :|: :: :||
Qy    426 AGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAE 466


RESULT    5
ENTRY            S41819     #type complete
TITLE            nucleoporin p62 - human
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
                   10-Sep-1997
ACCESSIONS       S41819; S43365
REFERENCE        S41819
   #authors      Carmo-Fonseca, M.; Kern, H.; Hurt, E.C.
   #journal      Eur. J. Cell Biol. (1991) 55:17-30
   #title        Human nucleoporin p62 and the essential yeast nuclear pore
                   protein NSP1 show sequence homology and a similar domain
                   organization.
   #accession    S41819
      ##molecule_type mRNA
      ##residues      1-522 ##label CAR
      ##cross-references EMBL:X58521
REFERENCE        S43365
   #authors      Hurt, E.C.
   #submission   submitted to the EMBL Data Library, March 1991
   #accession    S43365
      ##molecule_type mRNA
      ##residues      1-417,'RA',420-430,'Q',432-506,'V',508-522 ##label HUR
      ##cross-references EMBL:X58521; NID:g432653; PID:g432654
SUMMARY          #length 522  #molecular-weight 53268  #checksum 4396

  Query Match             3.0%;  Score 100;  DB 2;  Length 522;
  Best Local Similarity  31.6%;  Pred. No. 3.03e-01;
  Matches    18;  Conservative   19;  Mismatches  19;  Indels   1;  Gaps   1;

Db    258 LKAPGAASGTSTTTST-AATATATTTTSSSTTGFALNLKPLAPAGIPSNTAAAVTAP 313
          || :  | | ::  :   :::   |||:::||:| :: |||:|||  :: :: | :|
Qy    338 LKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPP 394


RESULT    6
ENTRY            S60024     #type complete
TITLE            bradykinin B1 receptor - rabbit
ORGANISM         #formal_name Oryctolagus cuniculus #common_name domestic
                   rabbit
DATE             24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change
                   10-Sep-1997
ACCESSIONS       S60024
```

```
REFERENCE       S60024
   #authors     MacNeil, T.; Bierilo, K.K.; Menke, J.G.; Hess, J.F.
   #journal     Biochim. Biophys. Acta (1995) 1264:223-228
   #title       Cloning and pharmacological characterization of a rabbit
                  bradykinin B(1) receptor.
   #accession   S60024
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-352 ##label MAC
      ##cross-references EMBL:U20507; NID:g1041820; PID:g1041821
KEYWORDS        G protein-coupled receptor; transmembrane protein
SUMMARY         #length 352  #molecular-weight 39503  #checksum 8863

  Query Match             2.9%;  Score 99;  DB 2;  Length 352;
  Best Local Similarity  25.0%;  Pred. No. 4.17e-01;
  Matches    25;  Conservative   35;  Mismatches  30;  Indels  10;  Gaps  10;

Db    216 AILFFNCHILA-SLRRRGERVPSRCG-GPRD-SKSTALILTLVASFLVCWAPYHFFAFLE 272
          ::|| :|: :| : |||   ||  |:  .|: ::  :|  :  ::: :|::    | | :
      210 GMLFCSCRDIACTERRRQTIVPV-CSYEERERPNCLSLQDSCKTNY-ICRSRLADF-FTN 266

      273 CLWQVHAIGGCFWEEFTDLGLQLSNFSAFVNSCLNPVIYV 312
          |  : :::: |: |:::|  | |: :|::: : : |  ||
Qy    267 CQPESRSVSNCLKENYAD-CL-LA-YSGLIGTVMTPN-YV 302


RESULT    7
ENTRY           I64089     #type complete
TITLE           protein-export protein - Haemophilus influenzae (strain Rd
                  KW20)
ORGANISM        #formal_name Haemophilus influenzae
DATE            18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
                  24-Oct-1997
ACCESSIONS      I64089
REFERENCE       A64000
   #authors     Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
                  Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
                  Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
                  FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
                  Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
                  J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
                  M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
                  D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
                  J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
                  Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
   #journal     Science (1995) 269:496-512
   #title       Whole-genome random sequencing and assembly of Haemophilus
                  influenzae Rd.
   #cross-references MUID:95350630
   #accession   I64089
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-169 ##label TIGR
      ##cross-references GB:U32758; GB:L42023; NID:g1573747; PID:g1573750;
                        TIGR:HI0743
FUNCTION
   #description  one of the proteins required for the normal export of
                   envelope proteins out of the cell cytoplasm; may be
                   involved in the initiation of the exporting process
CLASSIFICATION  #superfamily protein-export protein secB
KEYWORDS        protein export
SUMMARY         #length 169  #molecular-weight 19132  #checksum 8441

  Query Match             2.8%;  Score 95;  DB 2;  Length 169;
  Best Local Similarity  25.9%;  Pred. No. 1.44e+00;
  Matches    14;  Conservative   16;  Mismatches  23;  Indels   1;  Gaps   1;

Db     74 TLEDSGDVAFICEVKQAGVFTISGLEDVQMAHCLTSQCPNMLFPYARELVSNLV 127
          :|:||    :||  : |  ||    |  :::||    :: |::|:  |:: ::
Qy    245 SLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYS-GLIGTVM 297


RESULT    8
ENTRY           JC1121     #type complete
TITLE           leukocyte adhesion protein beta chain precursor - bovine
ALTERNATE_NAMES surface glycoprotein CD18
ORGANISM        #formal_name Bos primigenius taurus #common_name cattle
DATE            30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
                  05-Sep-1997
ACCESSIONS      JC1121
REFERENCE       JC1121
   #authors     Shuster, D.E.; Bosworth, B.T.; Kehrli Jr., M.E.
   #journal     Gene (1992) 114:267-271
   #title       Sequence of the bovine CD18-encoding cDNA: Comparison with
                  the human and murine glycoproteins.
   #cross-references MUID:92290287
   #accession   JC1121
      ##molecule_type mRNA
      ##residues      1-769 ##label SHU
      ##cross-references GB:M81233; NID:g162818; PID:g162819
COMMENT      The leukocyte adhesion proteins are noncovalently linked
               heterodimers of distinct alpha and identical beta chains. These
               structurally related glycoproteins mediate cell-adhesion
               reactions, and a deficiency of them is attributed to a genetic
               defect in the expression or structure of their common beta chain.
CLASSIFICATION  #superfamily integrin beta chain.
KEYWORDS        cell adhesion; cytoskeleton; duplication; glycoprotein;
                  heterodimer; leukocyte; pyroglutamic acid; transmembrane
                  protein
FEATURE
   1-22              #domain signal sequence #status predicted #label SIG\
   23-769            #product leukocyte adhesion protein beta chain #status
                       predicted #label MAT\
   23-670            #domain extracellular #status predicted #label EXT\
   445-631           #region cysteine-rich\
   701-723           #domain transmembrane #status predicted #label TMM\
   724-769           #domain intracellular #status predicted #label CYT\
   23                #modified_site pyrrolidone carboxylic acid (Gln) (in
                       mature form) #status predicted\
   50,116,254,501,
   642               #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY         #length 769  #molecular-weight 84399  #checksum 9562

  Query Match             2.8%;  Score 96;  DB 1;  Length 769;
  Best Local Similarity  30.8%;  Pred. No. 1.06e+00;
  Matches    24;  Conservative   20;  Mismatches  22;  Indels  12;  Gaps   8;

Db    477 IGKNCECQTQGRSSQELEGSCRKDNSS-II-CS-GLGDCICGQCVCHTS-----D-VPNK 527
          |:|  :|   :::  :|: :|:|  |: |  |: :::: :| :  || :     | || |
Qy    148 ISKGNNCLDAAKAC-NLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAK 206

Db    528 KIYGQ-FCECDNVNC-ER 543
            ||  || | :: | ||
Qy    207 HSYGMLFCSCRDIACTER 224


RESULT    9
ENTRY           A60385     #type complete
TITLE           monocyte surface antigen MS2 precursor - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change
                  13-Mar-1998
ACCESSIONS      A60385
REFERENCE       A60385
   #authors     Yoshida, S.; Setoguchi, M.; Higuchi, Y.; Akizuki, S.;
                  Yamamoto, S.
   #journal     Int. Immunol. (1990) 2:585-591
   #title       Molecular cloning of cDNA encoding MS2 antigen, a novel cell
                  surface antigen strongly expressed in murine monocytic
                  lineage.
   #accession   A60385
      ##molecule_type mRNA
      ##residues      1-826 ##label YOS
      ##cross-references EMBL:X13335
CLASSIFICATION  #superfamily disintegrin homology
```

```
KEYWORDS          glycoprotein; surface antigen; transmembrane protein
FEATURE
   1-14                #domain signal sequence #status predicted #label SIG\
   402-484             #domain disintegrin homology #label DIS\
   659-683             #domain transmembrane #status predicted #label TMM\
   330                 #active_site Glu #status predicted
SUMMARY           #length 826  #molecular-weight 89896  #checksum 2686

  Query Match             2.8%;  Score 96;  DB 2;  Length 826;
  Best Local Similarity  31.4%;  Pred. No. 1.06e+00;
  Matches    16;  Conservative   12;  Mismatches  17;  Indels   6;  Gaps   6;

Db    585 ELVLQGTKC-EEGKVC-MDGSCQDLR-VYRSENCSAKCNNHGVCNHKRECH 632
          | :  | :| : :| | :| :|   |  | :  |::   :: ||| :| ||
Qy    146 EHISKGNNCLDAAKACNLDDTCKKYRSAYITP-CTTS-MSNEVCN-RRKCH 193


RESULT   10
ENTRY             S33821     #type complete
TITLE             median body protein - Giardia lamblia
ORGANISM          #formal_name Giardia lamblia
DATE              06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
                    09-Sep-1997
ACCESSIONS        S33821
REFERENCE         S33821
   #authors       Marshall, J.; Holberton, D.V.
   #journal       J. Mol. Biol. (1993) 231:521-530
   #title         Sequence and structure of a new coiled coil protein from a
                    microtubule bundle in Giardia.
   #accession     S33821
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-857 ##label MAR
      ##cross-references EMBL:X64517; NID:g312670; PID:g312671
SUMMARY           #length 857  #molecular-weight 100583  #checksum 6805

  Query Match             2.8%;  Score 94;  DB 2;  Length 857;
  Best Local Similarity  33.3%;  Pred. No. 1.96e+00;
  Matches    15;  Conservative   16;  Mismatches  11;  Indels   3;  Gaps   3;

Db    482 KENSYNFDQLLEQKQQMRSDLNALREKAADYE-RVDRELRLKDKE 525
          || ::::   || |:: || ::||::|:  |: |  | :: |:|:
Qy     56 KETNFSLTSGLEAKDECRSAMEALKQKSL-YNCRCKRGMK-KEKN 98


RESULT   11
ENTRY             S72229     #type complete
TITLE             meiotic recombination protein mei-218 - fruit fly (Drosophila
                    melanogaster)
ORGANISM          #formal_name Drosophila melanogaster
DATE              04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
                    26-Feb-1998
ACCESSIONS        S72229
REFERENCE         S72229
   #authors       McKim, K.S.; Dahmus, J.B.; Hawley, R.S.
   #journal       Genetics (1996) 144:215-228
   #title         Cloning of the Drosophila melanogaster meiotic recombination
                    gene mei-218: a genetic and molecular analysis of interval
                    15E.
   #accession     S72229
      ##status         preliminary; nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues       1-1186 ##label MCK
      ##cross-references EMBL:U35631; NID:g1017733; PID:g1017734
GENETICS
   #gene          mei-218
   #introns       45/2; 558/3; 1003/2; 1081/1
SUMMARY           #length 1186  #molecular-weight 133551  #checksum 4925

  Query Match             2.8%;  Score 94;  DB 2;  Length 1186;
  Best Local Similarity  27.0%;  Pred. No. 1.96e+00;
  Matches    20;  Conservative   23;  Mismatches  25;  Indels   6;  Gaps   5;

Db    636 FDQQSGI-IWSNSSPQIPVEIQSNSVSQSVFRLASLPCTGVGTTNMDIFTSSPPPIQHMS 694
          |   | : :|  ::|  ||:  : : : : ||: : |  | : :: :| |   ||  ::
Qy    345 FGNGSDVTMW-QPAP--PVQTTTATTT-TAFRVKNKP-LGPAGSENEIPTHVLPPCANLQ 399

Db    695 SSRLPSSISGPTQL 708
          : :| |::||:|:|
Qy    400 AQKLKSNVSGSTHL 413


RESULT   12
ENTRY             S26728     #type fragment
TITLE             hypothetical protein 100-plus (rpoA2 3' region) -
                    Thermoplasma acidophilum (fragment)
ORGANISM          #formal_name Thermoplasma acidophilum
DATE              12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
                    09-Sep-1997
ACCESSIONS        S26728
REFERENCE         S26721
   #authors       Klenk, H.P.; Renner, O.; Schwass, V.; Zillig, W.
   #journal       Nucleic Acids Res. (1992) 20:5226
   #title         Nucleotide sequence of the genes encoding the subunits H, B,
                    A' and A'' of the DNA-dependent RNA polymerase and the
                    initiator tRNA from Thermoplasma acidophilum.
   #accession     S26728
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-100 ##label KLE
      ##cross-references EMBL:X68198; NID:g48089; PID:g48097
      ##experimental_source DSM 1728
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, September 1992
SUMMARY           #length 100  #checksum 201

  Query Match             2.7%;  Score 92;  DB 2;  Length 100;
  Best Local Similarity  31.5%;  Pred. No. 3.57e+00;
  Matches    17;  Conservative   15;  Mismatches  20;  Indels   2;  Gaps   2;

Db      4 GAYHICLNQS-FAGTG-ADRSFPASSEALLDPPEPELPRLPLKKFVNGPVLLSI 55
          |: |:||::| |:  | |  |   :: ::  ||   |: ||: ::   : |||:
Qy    409 GSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSV 462


RESULT   13
ENTRY             S58719     #type complete
TITLE             probable membrane protein YNL058c - yeast (Saccharomyces
                    cerevisiae)
ALTERNATE_NAMES   hypothetical protein N2433; hypothetical protein YNL1621
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              16-Feb-1996 #sequence_revision 23-Feb-1996 #text_change
                    12-Dec-1997
ACCESSIONS        S58719; S62986
REFERENCE         S58711
   #authors       Bergez, P.; Doignon, F.; Crouzet, M.
   #journal       Yeast (1995) 11:967-974
   #title         The sequence of a 44 420 bp fragment located on the left arm
                    of chromosome XIV from Saccharomyces cerevisiae.
   #accession     S58719
      ##status         nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues       1-316 ##label BER
      ##cross-references EMBL:U12141; NID:g1314216; PID:g994828
      ##note           the nucleotide sequence was submitted to the EMBL Data
                        Library, July 1994
REFERENCE         S62975
   #authors       Bergez, P.; Doignon, F.; Crouzet, M.
   #submission    submitted to the Protein Sequence Database, April 1996
   #accession     S62986
      ##molecule_type DNA
      ##residues       1-316 ##label BEW
      ##cross-references EMBL:Z71334; NID:g1301928; PID:e239896; PID:g1301929;
                          MIPS:YNL058c
      ##experimental_source strain S288C
GENETICS
```

```
   #map_position 14L
KEYWORDS          transmembrane protein
FEATURE
   95-111             #domain transmembrane #status predicted #label TMM
SUMMARY           #length 316  #molecular-weight 35046  #checksum 6749

  Query Match             2.7%;  Score 91;  DB 2;  Length 316;
  Best Local Similarity  28.6%;  Pred. No. 4.80e+00;
  Matches    18;  Conservative   22;  Mismatches  18;  Indels   5;  Gaps   5;

Db    153 INDEKSYMQDDWDSMSQ-LESSQYEDAASPFNPIQDPFTDNRRSL-FISPTLQ-VSQYEK 209
          :: ||: ::  | || | |:::: : ||::|:  ::|  |:: |||  :| | : |
Qy     93 MKKEKNCLRIYW-SMYQSLQGNDLLED-SPYEPVNSRLSDIFRAVPFISDVFQQVEHISK 150

Db    210 SHS 212
          :::
Qy    151 GNN 153


RESULT   14
ENTRY             A55356     #type complete
TITLE             urokinase-type plasminogen activator receptor membrane form
                    precursor - mouse
ALTERNATE_NAMES   urokinase-type plasminogen activator receptor 1
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
                    12-Dec-1997
ACCESSIONS        A55356; A41643
REFERENCE         A55356
   #authors       Suh, T.T.; Nerlov, C.; Dano, K.; Degen, J.L.
   #journal       J. Biol. Chem. (1994) 269:25992-25998
   #title         The murine urokinase-type plasminogen activator receptor
                    gene.
   #accession     A55356
      ##molecule_type DNA
      ##residues       1-327 ##label SUH
      ##cross-references GB:U12235; NID:g555809; PID:g555810
REFERENCE         A41643
   #authors       Kristensen, P.; Eriksen, J.; Blasi, F.; Dano, K.
   #journal       J. Cell Biol. (1991) 115:1763-1771
   #title         Two alternatively spliced mouse urokinase receptor mRNAs with
                    different histological localization in the gastrointestinal
                    tract.
   #cross-references MUID:92098587
   #accession     A41643
      ##status        preliminary
      ##molecule_type mRNA
      ##residues       1-327 ##label KRI
      ##cross-references GB:X62700; NID:g441462; PID:g441463
GENETICS
   #introns       20/1; 57/1; 105/1; 157/1; 202/1; 251/1
CLASSIFICATION    #superfamily urokinase-type plasminogen activator receptor;
                    Ly-6 homology
KEYWORDS          alternative splicing; duplication; glycoprotein;
                    phosphatidylinositol linkage
FEATURE
   24-108             #domain Ly-6 homology #label 1LY6\
   117-205            #domain Ly-6 homology #label 2LY6\
   213-301            #domain Ly-6 homology #label 3LY6\
   32,75,183,193,221,
   254,282            #binding_site carbohydrate (Asn) (covalent) #status
                        predicted
SUMMARY           #length 327  #molecular-weight 35428  #checksum 3430

  Query Match             2.7%;  Score 90;  DB 2;  Length 327;
  Best Local Similarity  24.2%;  Pred. No. 6.45e+00;
  Matches    23;  Conservative   22;  Mismatches  46;  Indels   4;  Gaps   4;

Db    192 CNYTHCNGGPVLDLQSFPPNGFQCYSCEGNNTLGCSSEEASLINC-RGPMNQCLVA-TGL 249
          |:|   :    | ||      : | |  ::    | :|  |: || :    :||:| :||
Qy    233 CSYEERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGL 292

Db    250 -DVLGNRSYTVRGC-ATASWCQGSHVADSFPTHLN 282
```

```
          :   :|   :  : |:||: |: :: :   |:
Qy    293 IGTVMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLK 327


RESULT   15
ENTRY             D70030     #type complete
TITLE             alkanal monooxygenase homolog yvbT - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    26-Feb-1998
ACCESSIONS        D70030
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #accession     D70030
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues       1-336 ##label KUN
      ##experimental_source strain 168
GENETICS
   #gene          yvbT
CLASSIFICATION    #superfamily ynbW protein
SUMMARY           #length 336  #molecular-weight 37101  #checksum 2405

  Query Match             2.7%;  Score 91;  DB 2;  Length 336;
  Best Local Similarity  40.0%;  Pred. No. 4.80e+00;
  Matches    16;  Conservative   12;  Mismatches  10;  Indels   2;  Gaps   2;

Db    132 NSGEDFPEQLEELRNYFKPSGNVRNQVRAIPGEGIDVPIW 171
          |||:|: : |  | |:|| :  ::| ::|: |:| || :|
Qy    317 NSGNDLEDCLKFL-NFFKDNTCLKNAIQAF-GNGSDVTMW 354


Search completed: Tue Jun 23 18:34:22 1998
Job time : 121 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

[redacted] on: Tue Jun 23 18:28:42 1998; MasPar time 16.07 Seconds
730.390 Million cell updates/sec
[redacted]ular output not generated.

Title: >US-08-866-354-4
Description: (1-468) from US08866354.pep
Perfect Score: 3384
Sequence: 1 MFLATLYFALPLLDLLMSAE.....PVLMLTALAALLSVSLAETS 468

Scoring table: PAM 150
Gap 11

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 47.975; Variance 67.645; scale 0.709

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3384	100.0	468	1	GDNR_RAT	GDNF RECEPTOR ALPHA PR	0.00e+00
2	3335	98.6	468	1	GDNR_MOUSE	GDNF RECEPTOR ALPHA PR	0.00e+00
3	3187	94.2	464	1	GDNR_HUMAN	GDNF RECEPTOR ALPHA PR	0.00e+00
4	2598	76.8	469	1	GDNR_CHICK	GDNF RECEPTOR ALPHA PR	0.00e+00
5	1556	46.0	464	1	NRTR_HUMAN	NEURTURIN RECEPTOR ALP	0.00e+00
6	1539	45.5	463	1	NRTR_MOUSE	NEURTURIN RECEPTOR ALP	0.00e+00
7	1511	44.7	465	1	NRTR_CHICK	NEURTURIN RECEPTOR ALP	0.00e+00
8	108	3.2	397	1	DP3B_MYCSM	DNA POLYMERASE III, BE	1.15e-03
9	102	3.0	249	1	ULA3_HCMVA	PROTEIN UL103.	1.28e-02
10	100	3.0	522	1	NU62_HUMAN	NUCLEAR PORE GLYCOPROT	2.80e-02
11	99	2.9	352	1	BRB1_RABIT	B1 BRADYKININ RECEPTOR	4.11e-02
12	95	2.8	169	1	SECB_HAEIN	PROTEIN-EXPORT PROTEIN	1.86e-01
13	96	2.8	504	1	GAG_HV1MA	GAG POLYPROTEIN (CONTA	1.28e-01
14	96	2.8	769	1	ITB2_BOVIN	CELL SURFACE ADHESION	1.28e-01
15	96	2.8	826	1	MS2_MOUSE	CELL SURFACE ANTIGEN M	1.28e-01
16	94	2.8	857	1	MEDB_GIALA	MEDIAN BODY PROTEIN.	2.69e-01
17	92	2.7	100	1	YRP4_THEAC	HYPOTHETICAL PROTEIN I	5.57e-01
18	91	2.7	316	1	YNF8_YEAST	HYPOTHETICAL 35.0 KD P	7.97e-01
19	92	2.7	320	1	KITH_PRVN3	THYMIDINE KINASE (EC 2	5.57e-01
20	90	2.7	327	1	UPAR_MOUSE	UROKINASE PLASMINOGEN	1.14e+00
21	91	2.7	388	1	GAG_HV1W2	GAG POLYPROTEIN (CONTA	7.97e-01
22	91	2.7	469	1	BF2_HUMAN	TRANSCRIPTION FACTOR B	7.97e-01
23	90	2.7	475	1	SIM1_YEAST	SIM1 PROTEIN.	1.14e+00
24	91	2.7	503	1	GAG_HV1JR	GAG POLYPROTEIN (CONTA	7.97e-01
25	93	2.7	567	1	CHI3_CANAL	CHITINASE 3 PRECURSOR	3.88e-01
26	93	2.7	600	1	SP96_DICDI	SPORE COAT PROTEIN SP9	3.88e-01
27	92	2.7	688	1	C1S_HUMAN	COMPLEMENT C1S COMPONE	5.57e-01
28	93	2.7	695	1	CASP_MESAU	CALCIUM-DEPENDENT SERI	3.88e-01
29	93	2.7	798	1	HEPA_HCMVA	DNA HELICASE/PRIMASE C	3.88e-01
30	89	2.6	247	1	SUMT_PSEFL	UROPORPHYRIN-III C-MET	1.62e+00
31	88	2.6	378	1	LEUK_RAT	LEUKOSIALIN PRECURSOR	2.29e+00
32	88	2.6	415	1	CBPB_RAT	CARBOXYPEPTIDASE B PRE	2.29e+00
33	89	2.6	491	1	CC1_YEAST	CELL DIVISION CONTROL	1.62e+00
34	88	2.6	506	1	GAG_HV1MN	GAG POLYPROTEIN (CONTA	2.29e+00
35	87	2.6	514	1	GUXC_FUSOX	PUTATIVE EXOGLUCANASE	3.22e+00
36	88	2.6	520	1	GAG_HV2SB	GAG POLYPROTEIN (CORE	2.29e+00
37	87	2.6	525	1	NU62_RAT	NUCLEAR PORE GLYCOPROT	3.22e+00
38	89	2.6	563	1	ENV_BAEVM	ENV POLYPROTEIN PRECUR	1.62e+00
39	88	2.6	743	1	STB5_YEAST	STB5 PROTEIN.	2.29e+00
40	88	2.6	769	1	ITB2_PIG	CELL SURFACE ADHESION	2.29e+00
41	89	2.6	770	1	ITB2_MOUSE	CELL SURFACE ADHESION	1.62e+00
42	88	2.6	1188	1	PPSA_METJA	PROBABLE PHOSPHOENOLPY	2.29e+00
43	88	2.6	1409	1	HAP1_HAEIN	ADHESION AND PENETRATI	2.29e+00
44	88	2.6	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	2.29e+00
45	89	2.6	2524	1	NOTC_XENLA	NEUROGENIC LOCUS NOTCH	1.62e+00

ALIGNMENTS

```
RESULT    1
ID   GDNR_RAT      STANDARD;      PRT;   468 AA.
AC   Q62997;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE   NEUROTROPHIC FACTOR RECEPTOR 1).
GN   GDNFRA OR TRNR1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 96270513.
RA   JING S., WEN D., YU Y., HOLST P.L., LUO Y., FANG M., TAMIR R.,
RA   ANTONIO L., HU Z., CUPPLES R., LOUIS J.-C., HU S., ALTROCK B.W.,
RA   FOX G.M.;
RL   CELL 85:1113-1124(1996).
CC   -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC       WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, BRAIN AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; U59486; G1399863; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GDNF RECEPTOR ALPHA.
FT   PROPEP        ?    468       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   DOMAIN      362    369       POLY-THR.
FT   CARBOHYD     59     59       POTENTIAL.
FT   CARBOHYD    347    347       POTENTIAL.
FT   CARBOHYD    406    406       POTENTIAL.
SQ   SEQUENCE   468 AA;  51649 MW;  6A7A2B2A CRC32;

  Query Match          100.0%;  Score 3384;  DB 1;  Length 468;
  Best Local Similarity 100.0%;  Pred. No. 0.00e+00;
  Matches   468;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
```

```
Db      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

       361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    2
ID   GDNR_MOUSE     STANDARD;      PRT;   468 AA.
AC   P97785;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE   NEUROTROPHIC FACTOR RECEPTOR 1).
GN   GDNFRA OR TRNR1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=DORSAL ROOT GANGLION;
RA   WATABE K.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC       WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; AB000800; G1816442; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GDNF RECEPTOR ALPHA.
FT   PROPEP        ?    468       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   DOMAIN      362    369       POLY-THR.
FT   CARBOHYD     59     59       POTENTIAL.
FT   CARBOHYD    347    347       POTENTIAL.
FT   CARBOHYD    406    406       POTENTIAL.
SQ   SEQUENCE   468 AA;  51782 MW;  6C64C182 CRC32;

  Query Match           98.6%;  Score 3335;  DB 1;  Length 468;
  Best Local Similarity  97.4%;  Pred. No. 0.00e+00;
  Matches   456;  Conservative   10;  Mismatches   2;  Indels   0;  Gaps   0;

Db       1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 240
           ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Qy     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db     241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
           ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db     301 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
           |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db     361 QTTTAMTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 420
           ||||| |||||:|||||||||||||||||||||||||||||||||||||||||||||:|:
Qy     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db     421 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVSLAETS 468
           |||||||||||||||||||||:|||||:::|||||||||||||||||
Qy     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    3
ID   GDNR_HUMAN     STANDARD;      PRT;   464 AA.
AC   P56159;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE   NEUROTROPHIC FACTOR RECEPTOR 1).
GN   GDNFRA OR TRNR1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SUBSTANTIA NIGRA;
RX   MEDLINE; 96270513.
RA   JING S., WEN D., YU Y., HOLST P.L., LUO Y., FANG M., TAMIR R.,
RA   ANTONIO L., HU Z., CUPPLES R., LOUIS J.-C., HU S., ALTROCK B.W.,
RA   FOX G.M.;
RL   CELL 85:1113-1124(1996).
CC   -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC       WITH THE DISULFIDE-LINKED GDNF DIMER AND 2 MOLECULES OF RET (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   MIM; 601496; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GDNF RECEPTOR ALPHA.
FT   PROPEP        ?    464       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   DOMAIN      361    368       POLY-THR.
FT   CARBOHYD     59     59       POTENTIAL.
FT   CARBOHYD    346    346       POTENTIAL.
FT   CARBOHYD    405    405       POTENTIAL.
SQ   SEQUENCE   464 AA;  51291 MW;  2C8C3574 CRC32;

  Query Match           94.2%;  Score 3187;  DB 1;  Length 464;
```

```
  Best Local Similarity  93.1%;  Pred. No. 0.00e+00;
  Matches   431;  Conservative   26;  Mismatches   4;  Indels   2;  Gaps   2;

Db       1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
           |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy       1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
           ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRVVPFISDVFQQVEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 180
           |||||||||||||| |||||||||||||:|||||||||||||||| ||||||||||||||
Qy     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db     181 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 240
           |:||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Qy     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

       241 PNCL-LQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 299
           |||| |||||||||||||||||||||||||||:||||||||||||||||||||||||||||
       241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db     300 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAPPV 359
           |:||||||||||||||||||||:|||||||||||||||||||||||||||:|||||||||
Qy     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db     360 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 419
           ||||||||||:|||||||||||||||||||||||||||||||||||||:||||:|::::
Qy     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db     420 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 461
             |:|| ||||||||||||||||:||:| ||::|||::|||::
Qy     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVS 463


RESULT    4
ID   GDNR_CHICK     STANDARD;      PRT;   469 AA.
AC   O13156;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GDNF RECEPTOR ALPHA PRECURSOR (GDNFR-ALPHA) (TGF-BETA RELATED
DE   NEUROTROPHIC FACTOR RECEPTOR 1).
GN   GDNFRA.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BUJ-BELLO A., ADU J., PINON L.G., HORTON A., THOMPSON J.,
RA   ROSENTHAL A., CHINCHETRU M., BUCHMAN V.L., DAVIES A.M.;
RL   NATURE 387:721-724(1997).
CC   -!- FUNCTION: RECEPTOR FOR GDNF. MEDIATES THE GDNF-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: 2 MOLECULES OF GDNFR-ALPHA ARE THOUGHT TO FORM A COMPLEX
CC       WITH THE DISULFIDE-LINKED GDNF DIMER AND WITH 2 MOLECULES OF RET
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; U90541; G2213803; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28      ?       GDNF RECEPTOR ALPHA.
FT   PROPEP        ?    469       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   DOMAIN      361    368       POLY-THR.
FT   CARBOHYD     62     62       POTENTIAL.
FT   CARBOHYD    163    163       POTENTIAL.
FT   CARBOHYD    346    346       POTENTIAL.
FT   CARBOHYD    405    405       POTENTIAL.

SQ   SEQUENCE   469 AA;  52043 MW;  C8D241C9 CRC32;

  Query Match           76.8%;  Score 2598;  DB 1;  Length 469;
  Best Local Similarity  78.1%;  Pred. No. 0.00e+00;
  Matches   367;  Conservative   49;  Mismatches   41;  Indels   13;  Gaps   9;

Db       1 MFLALLYLALPLADVLLSAEVSGLPGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKE 60
           |||| ||:|||| |:|:|||||   ||||||||||||||||||||||||||||||||||
Qy       1 MFLATLYFALPLLDLLMSAEVSG---GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKE 57

Db      61 SNFSRATGLEAKDECKSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLE 120
           :||| ::|||||||:|||||||||||||||||||||||||||||||||||||||||||||
Qy      58 TNFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLE 117

Db     121 DSPYEPVNSRLSDIFRLAPIVS-V-EPV--LSKGNNCLDAAKACNLNDTCKRFRSAYITP 176
           |||||||||||||||  |::| | : |  :|||||||||||||:||||::|||||||
Qy     118 DSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITP 177

Db     177 CTSSTSNEICNKRKCHKALRLFFDKVPPKHSYGMLFCSCRDVACTERRRQTIVPVCSYED 236
           ||:| |||:||:|||||||| ||||||:|||||||||||||:|||||||||||||||||:
Qy     178 CTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEE 237

Db     237 REKPNCLNLQESCKKNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVM 296
           ||:||||:||:||| |||||||||||||||||||:|||||||||||||||||||||||||
Qy     238 RERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVM 297

Db     297 TPNYIDSSSLSVAPWCDCSNSGNDIDECRKFLNFFQDNTCLKNAIQAFGNGTDVNVWQPI 356
           ||||:|||||||||||||||:::| |||||| |||||||||||||:|| :|||
Qy     298 TPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPA 357

Db     357 LPVQTTTATTTTASRLKNT--GSETTNNEIPTHNDSPACANLQAQKKRKSNESVDTELCL 414
           |||||||||||| |:||   |:  ::|||||   |:||||||  ||| |  | |||
Qy     358 PPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHV-LPPCANLQAQKL-KSNVSGSTHLCL 415

Db     415 NENAIGKDNTPGVSTSHISSENSFALPTSFYPSTPLILMTIALSLFLFLS 464
           ::: :||| :| | |||:: :  | |:   :| | :||  ||: :| :|
Qy     416 SDSDFGKDGLAGAS-SHITTKSMAAPPSCSLSSLP-VLMLTALAALLSVS 463


RESULT    5
ID   NRTR_HUMAN     STANDARD;      PRT;   464 AA.
AC   O00451;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (TGF-
DE   BETA RELATED NEUROTROPHIC FACTOR RECEPTOR 2) (GDNF RECEPTOR BETA)
DE   (GDNFR-BETA).
GN   GDNFRB OR TRNR2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97325791.
RA   BALOH R.H., TANSEY M.G., GOLDEN J.P., CREEDON D.J.,
RA   HEUCKEROTH R.O., KECK C.L., ZIMONJIC D.B., POPESCU N.C.,
RA   JOHNSON E.M., MILBRANDT J.;
RL   NEURON 18:793-802(1997).
CC   -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; AF002700; G2145080; -.
DR   MIM; 601956; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?      ?       TGF-BETA RELATED NEUROTROPHIC FACTOR
FT                                RECEPTOR 2.
FT   PROPEP        ?    464       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   CARBOHYD     52     52       POTENTIAL.
```

```
FT   CARBOHYD    357    357       POTENTIAL.
FT   CARBOHYD    413    413       POTENTIAL.
SQ   SEQUENCE   464 AA;  51558 MW;  3C74BBFB CRC32;

  Query Match           46.0%;  Score 1556;  DB 1;  Length 464;
  Best Local Similarity  48.4%;  Pred. No. 0.00e+00;
  Matches   224;  Conservative  101;  Mismatches 120;  Indels  18;  Gaps  15;

Db     12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy      2 FLATLYFALPLLDLLMSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEEFYEAS 126
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy    120 PYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy    180 TSMS-NEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
Qy    239 ERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
Qy    299 PNYVDSS-S-LSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQP 356

Db    367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
Qy    357 APPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLS 416

Db    424 E--LTTNIIPGSNKVIKPNSGPSRARPSAALTVLSVLMLKQAL 464
Qy    417 DSDFGKDGLAGASSHITTKS-MA-APPSCSLSSLPVLMLT-AL 456


RESULT    6
ID   NRTR_MOUSE     STANDARD;      PRT;   463 AA.
AC   O08842;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (TGF-
DE   BETA RELATED NEUROTROPHIC FACTOR RECEPTOR 2) (GDNF RECEPTOR BETA)
DE   (GDNFR-BETA).
GN   GDNFRB OR TRNR2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97325791.
RA   BALOH R.H., TANSEY M.G., GOLDEN J.P., CREEDON D.J.,
RA   HEUCKEROTH R.O., KECK C.L., ZIMONJIC D.B., POPESCU N.C.,
RA   JOHNSON E.M., MILBRANDT J.;
RL   NEURON 18:793-802(1997).
CC   -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; AF002701; G2145082; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?      ?       TGF-BETA RELATED NEUROTROPHIC FACTOR
FT                                RECEPTOR 2.
FT   PROPEP        ?    463       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
```

```
FT   CARBOHYD     52     52       POTENTIAL.
FT   CARBOHYD    357    357       POTENTIAL.
FT   CARBOHYD    413    413       POTENTIAL.
SQ   SEQUENCE   463 AA;  51598 MW;  0A2165C0 CRC32;

  Query Match           45.5%;  Score 1539;  DB 1;  Length 463;
  Best Local Similarity  49.1%;  Pred. No. 0.00e+00;
  Matches   214;  Conservative   96;  Mismatches 113;  Indels  13;  Gaps  11;

Db     37 QVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN-TM---LANK-ECQAALEVLQESPLY 91
Qy     26 RLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLY 85

Db     92 DCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEASPYEPVTSRLSDIFRLASIFSGTGAD 151
Qy     86 NCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQ 144

Db    152 PVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICNREISPTERCNRRKCHKALRQFFDRV 211
Qy    145 VEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTTSMS-NEVCNRRKCHKALRQFFDKV 203

Db    212 PSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDKEKPNCLDLRSLCRTDHLCRSRLADF 271
Qy    204 PAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADF 263

Db    272 HANCRASYRTITSCPADNYQACLGSYAGMIGFDMTPNYVDSNPTGIVVSPWCNCRGSGNM 331
Qy    264 FTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPNYVDSS-S-LSVAPWCDCSNSGND 321

Db    332 EEECEKFLKDFTENPCLRNAIQAFGNGTDVNM-SPKGPTFSATQAPRVE-KTPSLPDDLS 389
Qy    322 LEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPA 381

Db    390 DSTS-LGTSVITTCTSIQEQGLKANNSKELSMCFTELTTNISPGSKKVIKLYSGSCRARL 448
Qy    382 GSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPP 441

Db    449 STALTALPLLMVT-LA 463
Qy    442 SCSLSSLPVLMLTALA 457


RESULT    7
ID   NRTR_CHICK     STANDARD;      PRT;   465 AA.
AC   O13157;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NEURTURIN RECEPTOR ALPHA PRECURSOR (NTNR-ALPHA) (NRTNR-ALPHA) (GDNF
DE   RECEPTOR BETA) (GDNFR-BETA).
GN   GDNFRB.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BUJ-BELLO A., ADU J., PINON L.G., HORTON A., THOMPSON J.,
RA   ROSENTHAL A., CHINCHETRU M., BUCHMAN V.L., DAVIES A.M.;
RL   NATURE 387:721-724(1997).
CC   -!- FUNCTION: RECEPTOR FOR NEURTURIN. MEDIATES THE NRTN-INDUCED
CC       AUTOPHOSPHORYLATION AND ACTIVATION OF THE RET RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GDNFR FAMILY.
DR   EMBL; U90542; G2213805; -.
KW   RECEPTOR; GLYCOPROTEIN; GPI-ANCHOR; MEMBRANE; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?      ?       NEURTURIN RECEPTOR ALPHA.
FT   PROPEP        ?    465       HYDROPHOBIC, REMOVED DURING MATURATION
FT                                (POTENTIAL).
FT   CARBOHYD    355    355       POTENTIAL.
FT   CARBOHYD    387    387       POTENTIAL.
FT   CARBOHYD    412    412       POTENTIAL.
```

```
SQ   SEQUENCE   465 AA;  51908 MW;  5CA073E4 CRC32;

  Query Match           44.7%;  Score 1511;  DB 1;  Length 465;
  Best Local Similarity  46.6%;  Pred. No. 0.00e+00;
  Matches   203;  Conservative  102;  Mismatches 116;  Indels  15;  Gaps  11;

Db      38 VDCIRANKLCAAEGSCSSRYRTLRQCLAGRDRN-TM---LANK-ECQAALEVLQESPLYD 92
           :||::|:  |  | |||::|||||||:||:: | ::   |  | ||::|:| | : :||:
Qy      27 LDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNFSLTSGLEAKDECRSAMEALKQKSLYN 86

Db      93 CRCKRGMRKEIQCLQVYWSIHLGLAEGEEFYEASPYEPITSRLSDIFRLASIFSGM-DPA 151
           |||||||:||  ||::|||:  :| :|::: | |||||: |||||||  :::| : :
Qy      87 CRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDSPYEPVNSRLSDIFRAVPFISDVFQQV 145

Db     152 TN-SKSNHCLDAAKACNLNDNCKRLRSGYISTCSKEISATEHCSRRKCHKALRQFFDNVP 210
            : ||:|:||||||||||:| ||: ||:||: |:  :|  | |:|||||||||||||:||
Qy     146 EHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTTSMSN-EVCNRRKCHKALRQFFDKVP 204

Db     211 SEYTYRLLFCSCKDQACAEPRRQTIVPFCSYEDKEKPNCLDLRNVCRADHLCRSRLADFH 270
           :  :| :|||||:| ||:| |||||| ||||::|:|||| |:: |::: :|||||||
Qy     205 AKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERERPNCLSLQDSCKTNYICRSRLADFF 264

Db     271 ANCQASFQSLTSCPGDNYQACLGSYTGLIGFDMTPNYVDASTTSITISPWCSCKGSGNLE 330
           :|||:  :|:::|  :||  || :|:|||  |||||||:|  |::::||| |  |||
Qy     265 TNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPNYVDSS--SLSVAPWCDCSNSGNDL 322

Db     331 EECEKFLRDFTENPCLRNAIQAFGNGTDVNL-SPKNPSPPITMLPKVE-KSPALPDDIND 388
           |:| |||  | :| ||:||||||||||:|| :  |  |    |       :    |
Qy     323 EDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAG 382

Db     389 SNTMYDTSIITTCTSIQEHGQKLNKSKEQSLCYSETQLTTDTMPDQKTFVDQKAAGSRHR 448
           |:    | ::  |:::| :  | | |    || |::::  | ::   : :  |: ::
Qy     383 SENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDFGKDGLAGASSHITTKSMAAPPS 442

Db     449 AARILPAVPIVLLKLL 464
             :  |:::|:::|  |
Qy     443 CS--LSSLPVLMLTAL 456


RESULT    8
ID   DP3B_MYCSM     STANDARD;      PRT;   397 AA.
AC   P52851;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7).
GN   DNAN.
OS   MYCOBACTERIUM SMEGMATIS.
     PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
     [1]
     SEQUENCE FROM N.A.
RC   STRAIN=MC26;
RX   MEDLINE; 96310367.
RA   SALAZAR L., FSIHI H., DE ROSSI E., RICCARDI G., RIOS C.,
RA   COLE S.T., TAKIFF H.E.;
RL   MOL. MICROBIOL. 20:283-293(1996).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT
CC       IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-
CC       INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC       N PYROPHOSPHATE + DNA(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND
CC       THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE
CC       IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-
CC       STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,
CC       GAMMA, AND DELTA.
DR   EMBL; X92503; E208988; -.
KW   DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
SQ   SEQUENCE   397 AA;  41324 MW;  EFFB3064 CRC32;

  Query Match           3.2%;  Score 108;  DB 1;  Length 397;
  Best Local Similarity  25.0%;  Pred. No. 1.15e-03;
  Matches   27;  Conservative   27;  Mismatches  50;  Indels   4;  Gaps   4;

Db     172 ESVVLAATDRFRLAVRELTWVTTAGDVEAAVLVPAKTL-AEAAKAGTDGNQVHLALGSGA 230
           ::   ::|  ||:  : |    :  :: : || |   | |:  |:   |:  || | | :
Qy     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCL-SDS 418

Db     231 SVGKDGLLGIRSE-GKRSTTRLLDAEFPKFRQLLPAEHTAVATIGVAE 277
             ||||| |  |    :| :      :: :  |: :  :|: ::::||
Qy     419 DFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAE 466


RESULT    9
ID   ULA3_HCMVA     STANDARD;      PRT;   249 AA.
AC   P16734;
DT   01-AUG-1990 (REL. 15, CREATED)
DT   01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT   01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE   PROTEIN UL103.
GN   UL103.
OS   HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90269039.
RA   CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA   HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA   PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RL   CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC       EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.
DR   EMBL; X17403; G59711; -.
DR   PIR; S09868; S09868.
SQ   SEQUENCE   249 AA;  28636 MW;  239ADB2D CRC32;

  Query Match           3.0%;  Score 102;  DB 1;  Length 249;
  Best Local Similarity  38.3%;  Pred. No. 1.28e-02;
  Matches   18;  Conservative    7;  Mismatches  20;  Indels   2;  Gaps   2;

Db     106 CTVISCVENCNLTRKCLHDLLQYLDAVNVRESFGRLLHHSARRLICS 152
           ||     | ||  |||   | |::| |  : |:| :|  | | : |:
Qy     178 CTTSMSNEVCN-RRKCHKALRQFFDKVPAKHSYG-MLFCSCRDIACT 222


RESULT   10
ID   NU62_HUMAN     STANDARD;      PRT;   522 AA.
AC   P37198;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NUCLEAR PORE GLYCOPROTEIN P62 (NUCLEOPORIN P62).
GN   NUP62.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92007939.
RA   CARMO-FONSECA M., KERN H., HURT E.C.;
RL   EUR. J. CELL BIOL. 55:17-30(1991).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX.
CC       THE N-TERMINAL IS PROBABLY INVOLVED IN NUCLEOCYTOPLASMIC
CC       TRANSPORT. THE C-TERMINAL IS PROBABLY INVOLVED IN PROTEIN-PROTEIN
CC       INTERACTION VIA COILED-COIL FORMATION AND MAY FUNCTION IN
CC       ANCHORAGE OF P62 TO THE PORE COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CENTRAL REGION OF THE NUCLEAR PORE, WITHIN
CC       THE TRANSPORTER. DURING MITOTIC CELL DIVISION, IT ASSOCIATES WITH
CC       THE POLES OF THE MITOTIC SPINDLE.
CC   -!- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC   -!- PTM: O-GLYCOSYLATED. CONTAINS ABOUT 10 N-ACETYLGLUSAMINE SIDE
CC       CHAIN SITES PREDICTED FOR THE ENTIRE PROTEIN, AMONGST WHICH ONLY
CC       ONE IN THE C-TERMINAL.
```

```
CC   -!- SIMILARITY: HIGH, WITH RAT P62 NUCLEOPORIN. SIMILAR N-TERMINAL
CC       REPEAT ORGANIZATION WITH CENTRAL REPEATING UNITS OF YEAST NSP1.
DR   EMBL; X58521; G432654; -.
KW   NUCLEAR PROTEIN; TRANSPORT; GLYCOPROTEIN; HEPTAD REPEAT PATTERN;
KW   COILED COIL; REPEAT.
FT   DOMAIN        1    295       15 X 9 AA APPROXIMATE REPEATS.
FT   REPEAT        1      9       1.
FT   REPEAT       13     21       2.
FT   REPEAT       29     37       3.
FT   REPEAT       39     47       4.
FT   REPEAT       57     65       5.
FT   REPEAT       71     79       6.
FT   REPEAT       85     93       7.
FT   REPEAT      111    119       8.
FT   REPEAT      137    145       9.
FT   REPEAT      155    163       10.
FT   REPEAT      168    176       11.
FT   REPEAT      185    193       12.
     REPEAT      234    242       13.
     REPEAT      253    261       14.
     REPEAT      287    295       15.
FT   DOMAIN      330    505       COILED COIL.
FT   CARBOHYD    373    373       GLNAC (BY SIMILARITY).
FT   CARBOHYD    468    468       GLNAC (BY SIMILARITY).
SQ   SEQUENCE   522 AA;  53321 MW;  DCE0ACA2 CRC32;

  Query Match             3.0%;  Score 100;  DB 1;  Length 522;
  Best Local Similarity  31.6%;  Pred. No. 2.80e-02;
  Matches    18;  Conservative   19;  Mismatches  19;  Indels   1;  Gaps   1;

Db    258 LKAPGAASGTSTTTST-AATATATTTTSSSTTGFALNLKPLAPAGIPSNTAAAVTAP 313
          || :  | | ::  :   :::   |||:::||:| :: |||:|||  :: :: | :|
Qy    338 LKNAIQAFGNGSDVTMWQPAPPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPP 394


RESULT   11
ID   BRB1_RABIT     STANDARD;      PRT;   352 AA.
AC   P48748;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   B1 BRADYKININ RECEPTOR (BK-1 RECEPTOR).
GN   BDKRB1.
OS   ORYCTOLAGUS CUNICULUS (RABBIT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; LAGOMORPHA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=AORTA;
RX   MEDLINE; 96085127.
RA   MACNEIL T., BIERILO K.K., MENKE J.G., HESS J.F.;
RL   BIOCHIM. BIOPHYS. ACTA 1264:223-228(1995).
CC   -!- FUNCTION: THIS IS A RECEPTOR FOR BRADYKININ. COULD BE A FACTOR IN
CC       CHRONIC PAIN AND INFLAMMATION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- INDUCTION: THE B1 RECEPTOR IS SYNTHESIZED DE NOVO IN RESPONSE TO
CC       TISSUE INJURY.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; U20507; G1041821; -.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
KW   G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW   LIPOPROTEIN; PALMITATE.
FT   DOMAIN        1     39       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     40     63       1 (POTENTIAL).
FT   DOMAIN       64     72       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     73     97       2 (POTENTIAL).
FT   DOMAIN       98    110       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    111    132       3 (POTENTIAL).
FT   DOMAIN      133    154       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    177       4 (POTENTIAL).
FT   DOMAIN      178    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    225       5 (POTENTIAL).
FT   DOMAIN      226    246       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    247    271       6 (POTENTIAL).
FT   DOMAIN      272    290       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    291    314       7 (POTENTIAL).
FT   DOMAIN      315    352       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    109    188       BY SIMILARITY.
FT   LIPID       329    329       PALMITATE (POTENTIAL).
FT   CARBOHYD     13     13       POTENTIAL.
FT   CARBOHYD     21     21       POTENTIAL.
FT   CARBOHYD    184    184       POTENTIAL.
SQ   SEQUENCE   352 AA;  39503 MW;  70F06F62 CRC32;

  Query Match             2.9%;  Score 99;  DB 1;  Length 352;
  Best Local Similarity  25.0%;  Pred. No. 4.11e-02;
  Matches    25;  Conservative   35;  Mismatches  30;  Indels  10;  Gaps  10;

Db    216 AILFFNCHILA-SLRRRGERVPSRCG-GPRD-SKSTALILTLVASFLVCWAPYHFFAFLE 272
          ::|| :|: :| : |||   ||  |:   |: ::  :|  :  ::: :|::    | | :
Qy    210 GMLFCSCRDIACTERRRQTIVPV-CSYEERERPNCLSLQDSCKTNY-ICRSRLADF-FTN 266

Db    273 CLWQVHAIGGCFWEEFTDLGLQLSNFSAFVNSCLNPVIYV 312
          |  : :::: |: |:::|  | |: :|::: : : |  ||
Qy    267 CQPESRSVSNCLKENYAD-CL-LA-YSGLIGTVMTPN-YV 302


RESULT   12
ID   SECB_HAEIN     STANDARD;      PRT;   169 AA.
AC   P44853;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PROTEIN-EXPORT PROTEIN SECB.
GN   SECB OR HI0743.
OS   HAEMOPHILUS INFLUENZAE.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   PASTEURELLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RL   SCIENCE 269:496-512(1995).
CC   -!- FUNCTION: THIS IS ONE OF THE PROTEINS REQUIRED FOR THE NORMAL
CC       EXPORT OF ENVELOPE PROTEINS OUT OF THE CELL CYTOPLASM; IT MAY
CC       BE INVOLVED IN THE INITIATION OF THE EXPORTING PROCESS, BY BINDING
CC       TO THE NASCENT POLYPEPTIDE VIA A SIGNAL SEQUENCE, MAINTAINING
CC       A STABLE AND PRE-TRANSLOCATION CONFORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC       COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       (BY SIMILARITY).
DR   EMBL; U32758; G1573750; -.
DR   TIGR; HI0743; -.
KW   PROTEIN TRANSPORT; TRANSLOCATION.
SQ   SEQUENCE   169 AA;  19132 MW;  74BB8A0E CRC32;

  Query Match             2.8%;  Score 95;  DB 1;  Length 169;
  Best Local Similarity  25.9%;  Pred. No. 1.86e-01;
  Matches    14;  Conservative   16;  Mismatches  23;  Indels   1;  Gaps   1;

Db     74 TLEDSGDVAFICEVKQAGVFTISGLEDVQMAHCLTSQCPNMLFPYARELVSNLV 127
          :|:||    :||  : |  ||    |   :::||    :: |::|:  |:: ::
Qy    245 SLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYS-GLIGTVM 297


RESULT   13
ID   GAG_HV1MA      STANDARD;      PRT;   504 AA.
```

```
AC   P04594;
DT   13-AUG-1987 (REL. 05, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   GAG POLYPROTEIN (CONTAINS: CORE PROTEINS P17, P24, P2, P7, P1, P6).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (MAL ISOLATE) (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86245056.
RA   ALIZON M., WAIN-HOBSON S., MONTAGNIER L., SONIGO P.;
RL   CELL 46:63-74(1986).
CC   -!- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE
CC       ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL
CC       REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM
CC       MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY
CC       RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.
CC       GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY
CC       BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.
CC   -!- PTM: THE P24 PROTEIN IS PHOSPHORYLATED.
DR   EMBL; K03456; G328021; -.
DR   EMBL; X04415; G60229; -.
DR   EMBL; A07116; G492872; -.
DR   HSSP; P03348; 1HVN.
DR   HIV; K03456; GAG$MAL.
KW   AIDS; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION; PHOSPHORYLATION.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   LIPID         1      1       MYRISTATE (BY SIMILARITY).
FT   CHAIN         1    137       CORE PROTEIN P17 (MATRIX PROTEIN).
FT   CHAIN       138    368       CORE PROTEIN P24 (CORE ANTIGEN).
FT   CHAIN       369    383       CORE PROTEIN P2.
FT   CHAIN       384    437       CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).
FT   CHAIN       438    453       CORE PROTEIN P1.
FT   CHAIN       454    504       CORE PROTEIN P6.
SQ   SEQUENCE   504 AA;  56001 MW;  4500CD44 CRC32;

  Query Match             2.8%;  Score 96;  DB 1;  Length 504;
  Best Local Similarity  41.9%;  Pred. No. 1.28e-01;
  Matches    18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db      27 KYR-LKHLVWASRELERFALNPGLLETGEGCQQIMEQL-QSTL 67
           ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy      45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT   14
ID   ITB2_BOVIN     STANDARD;      PRT;   769 AA.
AC   P32592;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CELL SURFACE ADHESION GLYCOPROTEINS LFA-1, CR3 AND P150,95, BETA-
DE   SUBUNIT PRECURSOR (INTEGRIN BETA-2) (CD18 ANTIGEN).
GN   ITGB2 OR CD18.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92290287.
RA   SHUSTER D.E., BOSWORTH B.T., KEHRLI M.E. JR.;
RL   GENE 114:267-271(1992).
RN   [2]
RP   VARIANT LAD GLY-128.
RX   MEDLINE; 93028437.
RA   SHUSTER D.E., KEHRLI M.E. JR., ACKERMANN M.R., GILBERT R.O.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:9225-9229(1992).
CC   -!- FUNCTION: ASSOCIATES WITH ALPHA-L (LFA-1) TO INTERACT WITH ICAM-1,
CC       AND WITH ALPHA-M (MAC-1) OR ALPHA-X TO FORM THE RECEPTOR FOR THE
CC       IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT.
CC   -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-2 ASSOCIATES
CC       WITH DIFFERENT ALPHA CHAINS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE
CC       BONDS.
CC   -!- DISEASE: DEFECTS IN CD18 ARE THE CAUSE OF LEUKOCYTE ADHESION
CC       DEFICIENCY (LAD). THE MUTATION CAUSING LAD (GLY-128) IS PREVALENT
CC       AMONG HOLSTEIN CATTLE THROUGHOUT THE WORLD, PLACING THIS DISORDER
CC       AMONG THE MOST COMMON GENETIC DISEASES KNOWN IN ANIMAL
CC       AGRICULTURE. ALL CATTLE WITH THE MUTANT ALLELE ARE RELATED TO ONE
CC       BULL, WHO THROUGH THE USE OF ARTIFICIAL INSEMINATION SIRED MANY
CC       CALVES IN THE 1950S AND 1960S.
CC   -!- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF
CC       CELL-SURFACE RECEPTOR.
DR   EMBL; M81233; G162819; -.
DR   PIR; JC1121; JC1121.
DR   PROSITE; PS00243; INTEGRIN_BETA; 3.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_2.
KW   INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT;
KW   EXTRACELLULAR MATRIX; CYTOSKELETON; SIGNAL; DISEASE MUTATION.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    769       INTEGRIN BETA-2.
FT   DOMAIN       23    700       EXTRACELLULAR.
FT   TRANSMEM    701    723       POTENTIAL.
FT   DOMAIN      724    769       CYTOPLASMIC.
FT   DOMAIN      449    617       CYSTEINE-RICH REPEATS.
FT   REPEAT      449    496       I.
FT   REPEAT      497    540       II.
FT   REPEAT      541    581       III.
FT   REPEAT      582    617       IV.
FT   SITE        397    399       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD     50     50       POTENTIAL.
FT   CARBOHYD    116    116       POTENTIAL.
FT   CARBOHYD    254    254       POTENTIAL.
FT   CARBOHYD    501    501       POTENTIAL.
FT   CARBOHYD    642    642       POTENTIAL.
FT   VARIANT     128    128       D -> G (IN LAD).
SQ   SEQUENCE   769 AA;  84400 MW;  46F481EE CRC32;

  Query Match             2.8%;  Score 96;  DB 1;  Length 769;
  Best Local Similarity  30.8%;  Pred. No. 1.28e-01;
  Matches    24;  Conservative   20;  Mismatches  22;  Indels  12;  Gaps   8;

Db     477 IGKNCECQTQGRSSQELEGSCRKDNSS-II-CS-GLGDCICGQCVCHTS-----D-VPNK 527
           |:|  :|   :::  :|: :|:|  |: |  |: :::: :| :  || :     | || |
Qy     148 ISKGNNCLDAAKAC-NLDDTCKKYRSAYITPCTTSMSNEVCNRRKCHKALRQFFDKVPAK 206

Db     528 KIYGQ-FCECDNVNC-ER 543
             ||  || | :: | ||
Qy     207 HSYGMLFCSCRDIACTER 224


RESULT   15
ID   MS2_MOUSE      STANDARD;      PRT;   826 AA.
AC   Q05910;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   CELL SURFACE ANTIGEN MS2 PRECURSOR (EC 3.4.24.-) (MACROPHAGE CYSTEINE-
DE   RICH GLYCOPROTEIN).
GN   ADAM8 OR MS2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RA   YAMAMOTO S., YOSHIYAMA K., SETOGUCHI M., MATSUURA K., HIGUCHI Y.,
RA   AKIZUKI S.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=ICR;
```

```
RX   MEDLINE; 91197896.
RA   YOSHIDA S., SETOGUCHI M., HIGUCHI Y., AKIZUKI S., YAMAMOTO S.;
RL   INT. IMMUNOL. 2:585-591(1990).
CC   -!- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUKOCYTES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MACROPHAGES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
DR   EMBL; X13335; E264744; -.
DR   EMBL; X13335; E225615; -.
DR   MGD; MGI:107825; ADAM8.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW   TRANSMEMBRANE; GLYCOPROTEIN; ANTIGEN; ZINC; HYDROLASE;
KW   METALLOPROTEASE; SIGNAL.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    826       CELL SURFACE ANTIGEN MS2.
FT   DOMAIN       17    658       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    659    683       POTENTIAL.
FT   DOMAIN      684    826       CYTOPLASMIC (POTENTIAL).
FT   METAL       329    329       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    330    330       BY SIMILARITY.
FT   METAL       333    333       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       339    339       ZINC (CATALYTIC) (PROBABLE).
FT   CARBOHYD     89     89       POTENTIAL.
FT   CARBOHYD    260    260       POTENTIAL.
FT   CARBOHYD    431    431       POTENTIAL.
FT   CARBOHYD    614    614       POTENTIAL.
SQ   SEQUENCE   826 AA;  90046 MW;  7C26F36F CRC32;

  Query Match             2.8%;  Score 96;  DB 1;  Length 826;
  Best Local Similarity  31.4%;  Pred. No. 1.28e-01;
  Matches    16;  Conservative   12;  Mismatches  17;  Indels   6;  Gaps   6;

Db    585 ELVLQGTKC-EEGKVC-MDGSCQDLR-VYRSENCSAKCNNHGVCNHKRECH 632
          |  :  | :| : :| | :| :|   |  | :  |::   :: ||| :| ||
Qy    146 EHISKGNNCLDAAKACNLDDTCKKYRSAYITP-CTTS-MSNEVCN-RRKCH 193


Search completed: Tue Jun 23 18:29:33 1998
Job time : 51 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jun 23 18:29:51 1998; MasPar time 28.43 Seconds
693.273 Million cell updates/sec
Tabular output not generated.

Title: >US-08-866-354-4
Description: (1-468) from US08866354.pep
Perfect Score: 3384
Sequence: 1 MFLATLYFALPLLDLLMSAE.....PVLMLTALAALLSVSLAETS 468

Scoring table: PAM 150
Gap 11

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl5
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_vertebrate
13:sp_unclassified

Statistics: Mean 46.587; Variance 79.510; scale 0.586

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	3341	98.7	468	10	O35246	GDNF RECEPTOR ALPHA.	0.00e+00
2	3291	97.3	463	10	O35748	GDNFR-ALPHA/TRNR1-DELT	0.00e+00
3	3248	96.0	463	10	O35252	GDNF RECEPTOR BETA.	0.00e+00
4	3091	91.3	460	2	O15507	RET LIGAND 1.	0.00e+00
5	1562	46.2	464	10	O35977	GLIAL CELL LINE-DERIVE	0.00e+00
6	1556	46.0	464	2	O15316	GLIAL CELL LINE-DERIVE	0.00e+00
7	1556	46.0	464	2	O15328	RET LIGAND 2.	0.00e+00
8	696	20.6	397	10	O35118	GFRALPHA-3.	4.86e-131
9	696	20.6	397	10	O35325	GLIAL CELL LINE-DERIVE	4.86e-131
10	100	3.0	129	11	Q97845	MATRIX PROTEIN P17 (FR	2.57e-01
11	100	3.0	130	11	O36786	MA-P17 (FRAGMENT).	2.57e-01
12	100	3.0	130	11	O36808	MA-P17 (FRAGMENT).	2.57e-01
13	100	3.0	133	11	Q97725	MATRIX PROTEIN P17 (FR	2.57e-01
14	101	3.0	262	9	Q50381	DNAN PROTEIN (FRAGMENT	1.85e-01
15	103	3.0	399	9	Q50790	ORIGIN OF REPLICATION	9.44e-02
16	100	3.0	486	11	Q77804	GAG PROTEIN.	2.57e-01
17	103	3.0	492	11	Q80624	GAG PROTEIN.	9.44e-02
18	98	2.9	130	11	O36772	MA-P17 (FRAGMENT).	4.95e-01
19	99	2.9	243	11	P89796	GAG PROTEIN (FRAGMENT)	3.57e-01
20	98	2.9	359	2	O15268	SKAP55 PROTEIN.	4.95e-01
21	98	2.9	1127	3	Q94248	CODED FOR BY C. ELEGAN	4.95e-01
22	97	2.9	1217	9	Q45440	MNXG.	6.85e-01
23	96	2.8	70	11	P89780	GAG PROTEIN (FRAGMENT)	9.44e-01
24	95	2.8	129	11	O36910	MA-P17 (FRAGMENT).	1.30e+00
25	96	2.8	130	11	O36773	MA-P17 (FRAGMENT).	9.44e-01
26	96	2.8	130	11	Q76946	GAG PROTEIN (FRAGMENT)	9.44e-01
27	96	2.8	130	11	Q78625	GAG PROTEIN (FRAGMENT)	9.44e-01
28	96	2.8	130	11	Q76944	GAG PROTEIN (FRAGMENT)	9.44e-01
29	95	2.8	130	11	Q75884	GAG PROTEIN (FRAGMENT)	1.30e+00
30	95	2.8	130	11	Q75891	GAG PROTEIN (FRAGMENT)	1.30e+00
31	95	2.8	130	11	Q75890	GAG PROTEIN (FRAGMENT)	1.30e+00
32	95	2.8	130	11	Q75868	GAG PROTEIN (FRAGMENT)	1.30e+00
33	95	2.8	130	11	Q75866	GAG PROTEIN (FRAGMENT)	1.30e+00
34	95	2.8	130	11	Q75873	GAG PROTEIN (FRAGMENT)	1.30e+00
35	95	2.8	130	11	Q75870	GAG PROTEIN (FRAGMENT)	1.30e+00
36	95	2.8	130	11	Q75883	GAG PROTEIN (FRAGMENT)	1.30e+00
37	95	2.8	130	11	Q75864	GAG PROTEIN (FRAGMENT)	1.30e+00
38	95	2.8	130	11	Q75871	GAG PROTEIN (FRAGMENT)	1.30e+00
39	95	2.8	130	11	Q75867	GAG PROTEIN (FRAGMENT)	1.30e+00
40	96	2.8	133	11	Q97724	MATRIX PROTEIN P17 (FR	9.44e-01
41	95	2.8	242	11	Q79070	GAG PROTEIN.	1.30e+00
42	96	2.8	304	11	Q77701	GAG PROTEIN.	9.44e-01
43	96	2.8	492	11	Q79351	GAG PROTEIN.	9.44e-01
44	95	2.8	492	11	Q74452	GAG POLYPROTEIN.	1.30e+00
45	95	2.8	623	8	Q39204	BHLH PROTEIN.	1.30e+00

ALIGNMENTS

RESULT 1
ID O35246 PRELIMINARY; PRT; 468 AA.
AC O35246;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE GDNF RECEPTOR ALPHA.
GN GDNFR-ALPHA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57; TISSUE=LIVER;
RA DEY B.K., WONG Y.W., TOO H.P.;
RL NEUROREPORT 9:0-0(0001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57; TISSUE=LIVER;
RA DEY B.K., WONG Y.W., TOO H.P.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AF014117; G2624961; -.
SQ SEQUENCE 468 AA; 51752 MW; AFDCE6A1 CRC32;

Query Match 98.7%; Score 3341; DB 10; Length 468;
Best Local Similarity 97.6%; Pred. No. 0.00e+00;
Matches 457; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

```
Db      1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 240
          ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Qy    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240
```

```
Db    241 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db    301 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db    361 QTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 420
          ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||:|:
Qy    361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db    421 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVSLAETS 468
          |||||||||||||||||||:|||||||:::|||||||||||||||||
Qy    421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    2
ID    O35748         PRELIMINARY;    PRT;    463 AA.
AC    O35748;
DT    01-JAN-1998 (TREMBLREL. 05, CREATED)
DT    01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT    01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE    GDNFR-ALPHA/TRNR1-DELTA PROTEIN.
OS    RATTUS NORVEGICUS (RAT).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC    EUTHERIA; RODENTIA.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=HANNOVER;
RA    ZHONG J., ANNIES M., HEUMANN R.;
RL    SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AJ002072; E1154274; -.
SQ    SEQUENCE   463 AA;  51032 MW;  93277F91 CRC32;

  Query Match           97.3%;  Score 3291;  DB 10;  Length 463;
  Best Local Similarity  98.9%;  Pred. No. 0.00e+00;
  Matches   463;  Conservative    0;  Mismatches   0;  Indels   5;  Gaps   1;

Db      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFIS-----VEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 175
          ||||||||||||||||||     |||||||||||||||||||||||||||||||||||||
Qy    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db    176 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db    236 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db    296 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db    356 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db    416 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 463
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    3
ID    O35252         PRELIMINARY;    PRT;    463 AA.
AC    O35252;
DT    01-JAN-1998 (TREMBLREL. 05, CREATED)
DT    01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT    01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE    GDNF RECEPTOR BETA.
GN    GDNFR-BETA.
OS    MUS MUSCULUS (MOUSE).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC    EUTHERIA; RODENTIA.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=C57;
RA    DEY B.K., WONG Y.W., TOO H.P.;
RL    NEUROREPORT 9:0-0(0001).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=C57;
RA    DEY B.K., WONG Y.W., TOO H.P.;
RL    SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AF015172; G2624963; -.
SQ    SEQUENCE   463 AA;  51134 MW;  910EF17F CRC32;

  Query Match           96.0%;  Score 3248;  DB 10;  Length 463;
  Best Local Similarity  96.6%;  Pred. No. 0.00e+00;
  Matches   452;  Conservative   10;  Mismatches   1;  Indels   5;  Gaps   1;

Db      1 MFLATLYFVLPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db    121 YEPVNSRLSDIFRAVPFIS-----VEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 175
          ||||||||||||||||||     |||||||||||||||||||||||||||||||||||||
Qy    121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db    176 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDVACTERRRQTIVPVCSYEERER 235
          ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Qy    181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db    236 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 295
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db    296 YIDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 355
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db    356 QTTTATTTTAFRIKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDNDY 415
          ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||:|:
Qy    361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db    416 GKDGLAGASSHITTKSMAAPPSCGLSSLPVMVFTALAALLSVSLAETS 463
          |||||||||||||||||||||:||||||:::|||||||||||||||||
Qy    421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAETS 468


RESULT    4
ID    O15507         PRELIMINARY;    PRT;    460 AA.
AC    O15507;
DT    01-JAN-1998 (TREMBLREL. 05, CREATED)
DT    01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT    01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE    RET LIGAND 1.
GN    RETL1.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC    EUTHERIA; PRIMATES.
RN    [1]
RP    SEQUENCE FROM N.A.
```

```
RC   TISSUE=KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SUBSTANTIA NIGRA;
RA   HISHIKI T., KONDOH K., ICHIMIYA S., NIMURA Y., SEKI N., OZAKI T.,
RA   SAKIYAMA S., TAKAHASHI H., OHNUMA N., TANABE M., FUJIMURA S.,
RA   NAKAGAWARA A.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
     EMBL; U97144; G2282026; -.
     EMBL; U95847; G2459742; -.
     SEQUENCE   460 AA;  50838 MW;  022FFECA CRC32;

  Query Match           91.3%;  Score 3091;  DB 2;  Length 460;
  Best Local Similarity  91.8%;  Pred. No. 0.00e+00;
  Matches   425;  Conservative   27;  Mismatches   5;  Indels   6;  Gaps   2;

Db       1 MFLATLYFALPLLDLLLSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60
Qy       1 MFLATLYFALPLLDLLMSAEVSGGDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      61 SLASGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQGNDLLEDSP 120

Db     121 YEPVNSRLSDIFRVVPFIS-----VEHIPKGNNCLDAAKACNLDDICKKYRSAYITPCTT 175
Qy     121 YEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCTT 180

Db     176 SVSNDVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEEREK 235
Qy     181 SMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEERER 240

Db     236 PNCLNLQDSCKTNYICRSRLADFFTNCQPESRSVSSCLKENYADCLLAYSGLIGTVMTPN 295
Qy     241 PNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMTPN 300

Db     296 YIDSSSLSVAPWCDCSNSGNDLEECLKFLNFFKDNTCLKNAIQAFGNGSDVTVWQPAFPV 355
Qy     301 YVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQPAPPV 360

Db     356 QTTTATTTTALRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGNTHLCISNGNY 415
Qy     361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLSDSDF 420

Db     416 EKEGL-GASSHITTKSMAAPPSCGLSPLLVLVVTALSTLLSLT 457
Qy     421 GKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVS 463


RESULT    5
ID   O35977         PRELIMINARY;   PRT;   464 AA.
AC   O35977;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR RECEPTOR-BETA.
GN   GDNFR-BETA OR RETL2.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   TRUPP M., RAYNOSCHEK C., IBANEZ C.F.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN/KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN/KIDNEY;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF005226; G2232252; -.
DR   EMBL; U97143; G2282024; -.
SQ   SEQUENCE   464 AA;  51668 MW;  B1A2BD11 CRC32;

  Query Match           46.2%;  Score 1562;  DB 10;  Length 464;
  Best Local Similarity  47.9%;  Pred. No. 0.00e+00;
  Matches   221;  Conservative   100;  Mismatches 126;  Indels  14;  Gaps  12;

Db      12 FLDETLRSLASPSSLQGSELHGWRPQVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy       2 FLATLYFALPLLDLLMSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db      71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEAS 126
Qy      61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db     127 PYEPVTSRLSDIFRLASIFSGTGTDPAVSTKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy     120 PYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCT 179

Db     187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy     180 TSMS-NEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db     247 EKPNCLDLRSLCRTDHLCRSRLADFHANCRASYRTITSCPADNYQACLGSYAGMIGFDMT 306
Qy     239 ERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMT 298

Db     307 PNYVDSNPTGIVVSPWCNCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNMSPK 366
Qy     299 PNYVDSS-S-LSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQP 356

Db     367 GPSLPATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSIQEQGLKANNSKELSMCFT 423
Qy     357 APPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLS 416

Db     424 ELTTNISPGSKKVIKLNSGSSRARLSAALTALPLLMLTLAL 464
Qy     417 DSDFGKDGLAGASSHITTKSMAAPPSCSLSSLPVLMLT-AL 456


RESULT    6
ID   O15316         PRELIMINARY;   PRT;   464 AA.
AC   O15316;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR RECEPTOR BETA.
GN   GDNFR-BETA.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL BRAIN;
RA   WARTIOVAARA K., SUVANTO P., HORELLI-KUITUNEN N., LINDAHL M.,
RA   MOSHNYAKOV M., AIRAKSINEN M.S., PALOTIE A., SARIOLA H., SAARMA M.;
```

```
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U93703; G2228737; -.
SQ   SEQUENCE   464 AA;  51530 MW;  DE80D543 CRC32;

  Query Match           46.0%;  Score 1556;  DB 2;  Length 464;
  Best Local Similarity  48.4%;  Pred. No. 0.00e+00;
  Matches   224;  Conservative  101;  Mismatches 120;  Indels  18;  Gaps  15;

Db     12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy      2 FLATLYFALPLLDLLMSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEAS 126
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy    120 PYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy    180 TSMS-NEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
Qy    239 ERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
Qy    299 PNYVDSS-S-LSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQP 356

Db    367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
Qy    357 APPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLS 416

Db    424 E--LTTNIIPGSNKVIKPNSGPSRARPSAALTVLSVLMLKQAL 464
Qy    417 DSDFGKDGLAGASSHITTKS-MA-APPSCSLSSLPVLMLT-AL 456


RESULT    7
ID   O15328          PRELIMINARY;   PRT;   464 AA.
AC   O15328;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
     RET LIGAND 2.
     RETL2.
     HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:6238-6243(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   SANICOLA M., HESSION C.A., WORLEY D.S., CARMILLO P., EHRENFELS C.,
RA   WALUS L., ROBINSON S., JAWORSKI G., WEI H., TIZARD R., WHITTY A.,
RA   PEPINSKY R.B., CATE R.L.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U97145; G2282028; -.
SQ   SEQUENCE   464 AA;  51543 MW;  288A8BD8 CRC32;

  Query Match           46.0%;  Score 1556;  DB 2;  Length 464;
  Best Local Similarity  48.4%;  Pred. No. 0.00e+00;
  Matches   224;  Conservative  101;  Mismatches 120;  Indels  18;  Gaps  15;

Db     12 FLDETLRSLASPSSLQGPELHGWRPPVDCVRANELCAAESNCSSRYRTLRQCLAGRDRN- 70
Qy      2 FLATLYFALPLLDLLMSAEVSG-GDRLDCVKASDQCLKEQSCSTKYRTLRQCVAGKETNF 60

Db     71 TM---LANK-ECQAALEVLQESPLYDCRCKRGMKKELQCLQIYWSIHLGLTEGEEFYEAS 126
Qy     61 SLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSL-QGNDLLEDS 119

Db    127 PYEPVTSRLSDIFRLASIFSGTGADPVVSAKSNHCLDAAKACNLNDNCKKLRSSYISICN 186
Qy    120 PYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITPCT 179

Db    187 REISPTERCNRRKCHKALRQFFDRVPSEYTYRMLFCSCQDQACAERRRQTILPSCSYEDK 246
Qy    180 TSMS-NEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCRDIACTERRRQTIVPVCSYEER 238

Db    247 EKPNCLDLRGVCRTDHLCRSRLADFHANCRASYQTVTSCPADNYQACLGSYAGMIGFDMT 306
Qy    239 ERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGTVMT 298

Db    307 PNYVDSSPTGIVVSPWCSCRGSGNMEEECEKFLRDFTENPCLRNAIQAFGNGTDVNVSPK 366
Qy    299 PNYVDSS-S-LSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQAFGNGSDVTMWQP 356

Db    367 GPSFQATQAPRV-E-KTPSLPDDLSDSTS-LGTSVITTCTSVQEQGLKANNSKELSMCFT 423
Qy    357 APPVQTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGSTHLCLS 416

Db    424 E--LTTNIIPGSNKVIKPNSGPSRARPSAALTVLSVLMLKLAL 464
Qy    417 DSDFGKDGLAGASSHITTKS-MA-APPSCSLSSLPVLMLT-AL 456


RESULT    8
ID   O35118          PRELIMINARY;   PRT;   397 AA.
AC   O35118;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GFRALPHA-3.
GN   GFRALPHA-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL6; TISSUE=HEART;
RA   NOMOTO S., ITO S., YANG L., ADACHI K., KIUCHI K.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AB008833; D1024441; -.
SQ   SEQUENCE   397 AA;  44307 MW;  A80E0D24 CRC32;

  Query Match           20.6%;  Score 696;  DB 10;  Length 397;
  Best Local Similarity  34.4%;  Pred. No. 4.86e-131;
  Matches   120;  Conservative   71;  Mismatches 137;  Indels  21;  Gaps  17;

Db     11 LLMILLLVLSLWLPLGAGNSLATENRFVNSCTQARKKCEANPACKAAYQHLGSCTSSLSR 70
Qy      2 FLATLYFALPL-LDLLMSAEVSGGDRL-D-CVKASDQCLKEQSCSTKYRTLRQCVAGKET 58

Db     71 PLPLEES-AMSADCLEAAEQLRNSSLIDCRCHRRMKHQATCLDIYWTVHPARSLGDYELD 129
Qy     59 NFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQ-GNDLLE 117

Db    130 VSPYE--DTVTSKPWK-MN-LSKL-NMLK--PDSDLCLKFAMLCTLHDKCDRLRKAYGEA 182
Qy    118 DSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITP 177

Db    183 CS-GIR---CQRHLCLAQLRSFFEKAAESHAQGLLLCPCAPEDAGCGERRRNTIAPSCAL 238
Qy    178 CTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSC-R-DIACTERRRQTIVPVCSY 235

Db    239 PS-VTPNCLDLRSFCRADPLCRSRLMDFQTHCHPMDI-LGTCATEQ-SRCLRAYLGLIGT 295
```

```
              |||| |:  |::: :||||| || |:|:|    :: |  |  : || || |||||
Qy    236 EERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGT 295

Db    296 AMTPNFISKVNTTVALSCTCRGSGNLQDECEQLERSFSQNPCLVEAIAA 344
          ||||::   : :||  | |  |||  ::|  :   | :| || :|| |
Qy    296 VMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQA 344


RESULT    9
ID   O35325         PRELIMINARY;   PRT;   397 AA.
AC   O35325;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR FAMILY RECEPTOR ALPHA-3.
GN   GFRA-3.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   TRUPP M., RAYNOSCHEK C., IBANEZ C.F.;
RL   SUBMITTED (SEP-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF020305; G2429367; -.
SQ   SEQUENCE   397 AA;  44333 MW;  F0C0C841 CRC32;

  Query Match           20.6%;  Score 696;  DB 10;  Length 397;
  Best Local Similarity  34.4%;  Pred. No. 4.86e-131;
  Matches   120;  Conservative   71;  Mismatches 137;  Indels  21;  Gaps  16;

Db     11 LLMILLLVLSLWLPLGAGNSLATENRFVNSCTQARKKCEANPACKAAYQHLGSCTSSLSR 70
          :|  | : |:| | |  :  ::  :|: : |  |   |  : :| : |: |  | ::
Qy      2 FLATLYFALPL-LDLLMSAEVSGGDRL-D-CVKASDQCLKEQSCSTKYRTLRQCVAGKET 58

Db     71 PLPLEES-AMSADCLEAAEQLRNSSLIDCRCHRRMKHQATCLDIYWTVHPARSLGDYELD 129
          ::|  :     :|  | | |:  || :||| | || :  || |||::  :   |:  |:
Qy     59 NFSLTSGLEAKDECRSAMEALKQKSLYNCRCKRGMKKEKNCLRIYWSMYQSLQ-GNDLLE 117

Db    130 VSPYE--DTVTSKPWK-MN-LSKL-NMLK--PDSDLCLKFAMLCTLHDKCDRLRKAYGEA 182
          ||||  ::  |   : :  :| :   :   : :: ||  |: | | | | : | ||  :
Qy    118 DSPYEPVNSRLSDIFRAVPFISDVFQQVEHISKGNNCLDAAKACNLDDTCKKYRSAYITP 177

Db    183 CS-GIR---CQRHLCLAQLRSFFEKAAESHAQGLLLCPCPPEDAGCGERRRNTIAPSCAL 238
          |: ::    | |: |   || ||:| :  |: |:|:|:|   | :| |||| || | |:
Qy    178 CTTSMSNEVCNRRKCHKALRQFFDKVPAKHSYGMLFCSCR--DIACTERRRQTIVPVCSY 235

Db    239 PS-VTPNCLDLRSFCRADPLCRSRLMDFQTHCHPMDI-LGTCATEQ-SRCLRAYLGLIGT 295
          |||| |:  |::: :||||| || |:|:|    :: |  |  : || || |||||
Qy    236 EERERPNCLSLQDSCKTNYICRSRLADFFTNCQPESRSVSNCLKENYADCLLAYSGLIGT 295

Db    296 AMTPNFISKVNTTVALSCTCRGSGNLQDECEQLERSFSQNPCLVEAIAA 344
          ||||::   : :||  | |  |||  ::|  :   | :| || :|| |
Qy    296 VMTPNYVDSSSLSVAPWCDCSNSGNDLEDCLKFLNFFKDNTCLKNAIQA 344


RESULT   10
ID   Q97845         PRELIMINARY;   PRT;   129 AA.
AC   Q97845;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE   MATRIX PROTEIN P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BAB MNT;
RA   NARWA R., ROQUES P., COURPOTIN C., PARNETMATHIEU F., BOUSSIN F.,
RA   ROANE A., MARC D., LASFARGUES G., DORMONT D.;
RL   J. VIROL. 70:4474-4483(1996).
DR   EMBL; X99948; E258593; -.
KW   MATRIX PROTEIN.
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  14550 MW;  542AB924 CRC32;

  Query Match           3.0%;  Score 100;  DB 11;  Length 129;
  Best Local Similarity  41.9%;  Pred. No. 2.57e-01;
  Matches    18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db     28 KYR-LKHIVWASRELERFALNPGLLESAEGCQQLMEQL-QSTL 68
          ||| |:: | |::| : |:| :|| |: : |:  || | | :|
Qy     45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT   11
ID   O36786         PRELIMINARY;   PRT;   130 AA.
AC   O36786;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MA-P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LEIGH BROWN A.J., LOBIDEL D., WADE C.M., REBUS S., PHILLIPS N.,
RA   BRETTLE R.P., FRANCE A.J., LEEN C.S., MCMENAMIN J., MCMILLAN A.,
RA   MAW R.D., MULCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,
RA   PEUTHERER J.F.;
RL   VIROLOGY 235:166-177(1997).
DR   EMBL; AF014183; G2406708; -.
FT   NON_TER       1      1
FT   NON_TER     130    130
SQ   SEQUENCE   130 AA;  14584 MW;  428D9E75 CRC32;

  Query Match           3.0%;  Score 100;  DB 11;  Length 130;
  Best Local Similarity  41.9%;  Pred. No. 2.57e-01;
  Matches    18;  Conservative   11;  Mismatches   9;  Indels   5;  Gaps   5;

Db      5 KYK-LKHIVWASRELERFAVNPGLLETSEGCRQIMEQL-QPSL 45
          ||: |:: | |::| : |:: :|| |: : ||  || | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT   12
ID   O36808         PRELIMINARY;   PRT;   130 AA.
AC   O36808;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MA-P17 (FRAGMENT).
GN   GAG.
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   LEIGH BROWN A.J., LOBIDEL D., WADE C.M., REBUS S., PHILLIPS N.,
RA   BRETTLE R.P., FRANCE A.J., LEEN C.S., MCMENAMIN J., MCMILLAN A.,
RA   MAW R.D., MULCAHY F., ROBERTSON J.R., SANKAR K.N., SCOTT G., WYLD R.,
RA   PEUTHERER J.F.;
RL   VIROLOGY 235:166-177(1997).
DR   EMBL; AF014205; G2406752; -.
FT   NON_TER       1      1
FT   NON_TER     130    130
SQ   SEQUENCE   130 AA;  14527 MW;  7325723C CRC32;

  Query Match           3.0%;  Score 100;  DB 11;  Length 130;
  Best Local Similarity  39.5%;  Pred. No. 2.57e-01;
  Matches    17;  Conservative   12;  Mismatches   9;  Indels   5;  Gaps   5;

Db      5 KYK-LKHIVWASRELERFAVNPGLLETSEGCRQIMEQL-QSAL 45
```

```
          ||: |:: | |::| : |:: :|| |: : ||  ||.| | :|
Qy     45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT   13
ID   Q97725         PRELIMINARY;   PRT;   133 AA.
AC   Q97725;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE   MATRIX PROTEIN P17 (FRAGMENT).
OS   HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OC   VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;
OC   LENTIVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAL-M;
RA   NARWA R., ROQUES P., COURPOTIN C., PARNETMATHIEU F., BOUSSIN F.,
RA   ROANE A., MARC D., LASFARGUES G., DORMONT D.;
RL   J. VIROL. 70:4474-4483(1996).
RL   EMBL; Z79562; E261908; -.
KW   MATRIX PROTEIN.
FT   NON_TER     133    133
SQ   SEQUENCE   133 AA;  14724 MW;  E22E819C CRC32;

  Query Match             3.0%;  Score 100;  DB 11;  Length 133;
  Best Local Similarity  44.2%;  Pred. No. 2.57e-01;
  Matches    19;  Conservative    9;  Mismatches  10;  Indels   5;  Gaps   5;

Db     28 KYR-LKHIVWASRELERFALNPGLLEASEGCRQILGQL-QPSL 68
          ||| |:: | |::| : |:| :|| || : ||  :  | | ||
Qy     45 KYRTLRQCV-AGKETN-FSLTSGL-EAKDECRSAMEALKQKSL 84


RESULT   14
ID   Q50381         PRELIMINARY;   PRT;   262 AA.
AC   Q50381;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   DNAN PROTEIN (FRAGMENT).
GN   DNAN.
OS   MYCOBACTERIUM SMEGMATIS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2-MC155;
RX   MEDLINE; 96062237.
RA   RAJAGOPALAN M., QIN M.H., NASH D.R., MADIRAJU M.V.V.S.;
RL   J. BACTERIOL. 177:6527-6535(1995).
DR   EMBL; U17833; G1079500; -.
FT   NON_TER     262    262
SQ   SEQUENCE   262 AA;  27269 MW;  E0114FC9 CRC32;

  Query Match             3.0%;  Score 101;  DB 9;  Length 262;
  Best Local Similarity  30.6%;  Pred. No. 1.85e-01;
  Matches    22;  Conservative   14;  Mismatches  33;  Indels   3;  Gaps   3;

Db    172 ESVVLAATDRFRLAVRELTWVTTAGDVEAAVLVPAKTL-AEAAKAGTDGNQVHLALGSGA 230
          ::   ::|  ||:  : |    :  :: : || |   | |:  |:   |:  || | | :
Qy    361 QTTTATTTTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCL-SDS 418

Db    231 SVGKDGLLGIRS 242
            ||||| |  |
Qy    419 DFGKDGLAGASS 430


RESULT   15
ID   Q50790         PRELIMINARY;   PRT;   399 AA.
AC   Q50790;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE   ORIGIN OF REPLICATION AND GENES RNPA, RPMH, DNAA,
DE   DNAN, RECF.
GN   DNAN.
OS   MYCOBACTERIUM TUBERCULOSIS.
OC   PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   SALAZAR L., FSIHI H., DE ROSSI E., RICCARDI G., RIOS C.,
RA   COLE S.T., TAKIFF H.E.;
RL   MICROBIOLOGY 20:283-293(1996).
DR   EMBL; X92504; E208997; -.
SQ   SEQUENCE   399 AA;  41773 MW;  640BC5B7 CRC32;

  Query Match             3.0%;  Score 103;  DB 9;  Length 399;
  Best Local Similarity  27.7%;  Pred. No. 9.44e-02;
  Matches    28;  Conservative   27;  Mismatches  42;  Indels   4;  Gaps   4;

Db    178 ATDSRLAVRELEWSASSPDIEAAVLVPAKTL-AEAAKAGIGGSDVRLSLGTGPGVGKDGL 236
          :|  |:  : |  ::|  :| : || |   | |:  |: ::||  :| | : :  |||||
Qy    368 TTAFRVKNKPLGPAGSENEIPTHVLPPCANLQAQKLKSNVSGS-THLCL-SDSDFGKDGL 425

Db    237 LGISGN-GKRSTTRLLDAEFPKFRQLLPTEHTAVATMDVAE 276
            | |::   :| :      :: :  |: |  :|: :: :||
Qy    426 AGASSHITTKSMAAPPSCSLSSLPVLMLTALAALLSVSLAE 466


Search completed: Tue Jun 23 18:32:04 1998
Job time : 133 secs.
```